

84 872

STIC-Biot ch/ChemLib

From: Schnizer, Holly
Sent: Tuesday, January 21, 2003 4:44 PM
To: STIC-Biotech/ChemLib
Subject: seq. search for appl. no. 09/435,403

RECEIVED

JAN 22 2003

STIC-Biotech/ChemLib (STIC)

Please search the commercial and interference databases for the following

1) Arg Pro Asn Tyr Ser Arg Arg Leu Pro Lys Gly Val Lys His Leu Lys Asp Phe Pro Ile Leu Pro Gly Glu Ile
(this sequence is SEQ ID NO: 2 wherein Leu at position 3 is mutated to Asn)

Thank you.

Holly Schnizer
AU 1653
CM1-9E09
305-3722
mailbox: CM1-9B01

Point of Contact:
Toby Port
Technical Info. Specialist
CM1 6A04
703-308-3534

Searcher: _____
Phone: _____
Location: _____
Date Picked Up: 1/23
Date Completed: 1/23
Searcher Prep/Review: _____
Clerical: _____
Online time: _____

TYPE OF SEARCH:

NA Sequences: _____
AA Sequences: _____
Structures: _____
Bibliographic: _____
Litigation: _____
Full text: _____
Patent Family: _____
Other: _____

VENDOR/COST (where applic.)

STN: _____
DIALOG: _____
Questel/Orbit: _____
DRLink: _____
Lexis/Nexis: _____
Sequence Sys.: _____
WWW/Internet: _____
Other (specify): _____

OM protein - protein search, using sw model
Run on: January 23, 2003, 08:55:22 ; Search time 35 seconds
(without alignments)
95.179 Million cell updates/sec

Title: SEQ2-ASNAT3
Perfect score: 136
Sequence: 1 RPNYSRLPKGVKHLDFPILPGEI 25
Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5
Searched: 908470 seqs, 133250620 residues
Total number of hits satisfying chosen parameters: 908470
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_101002.*
1: /SIDS2/qcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDS2/qcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDS2/qcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDS2/qcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDS2/qcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDS2/qcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDS2/qcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDS2/qcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDS2/qcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDS2/qcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDS2/qcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDS2/qcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDS2/qcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDS2/qcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDS2/qcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDS2/qcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDS2/qcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDS2/qcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDS2/qcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDS2/qcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDS2/qcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDS2/qcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SIDS2/qcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES						
Result No.	Score	Query Match	Length	ID	Description	
1	127	93.4	368	16	AA73020	Human Factor-VIII
2	127	93.4	720	16	AA74088	Factor-VIII heavy
3	127	93.4	729	16	AA74089	Factor-VIII heavy
4	127	93.4	740	16	AA76961	Human Factor-VIII
5	127	93.4	740	16	AA76962	Human Factor-VIII
6	127	93.4	740	16	AA73021	Human Factor-VIII
7	127	93.4	740	16	AA74090	Factor-VIII heavy
8	127	93.4	740	16	AA76982	Human factor VIII
9	127	93.4	1383	18	AAW33227	Procoagulant-activ
10	127	93.4	1383	18	AAW33228	

11	127	93.4	1383	18	AAW33229	Procoagulant-activ
12	127	93.4	1424	9	AA780268	Modified factor VI
13	127	93.4	1424	10	AA791169	Sequence of 740 Ar
14	127	93.4	1424	22	AA748842	Mutant mature huma
15	127	93.4	1425	9	AA780267	Modified factor VI
16	127	93.4	1438	21	AA701262	B-domain deleted f
17	127	93.4	1440	12	AA712971	Factor VIII:SQ
18	127	93.4	1457	19	AAW46246	Human factor VIII
19	127	93.4	1457	19	AAW44372	Human Factor VIII
20	127	93.4	1457	20	AAV21675	Beta-domain delete
21	127	93.4	1459	22	AAE10827	Human factor VIII
22	127	93.4	1459	22	AAE10832	Human factor VIII
23	127	93.4	1459	22	AAE10833	Human factor VIII
24	127	93.4	1471	18	AAW23414	Human B-domain del
25	127	93.4	1471	22	AA767959	Amino acid sequenc
26	127	93.4	1516	9	AA780265	Modified factor VI
27	127	93.4	1661	18	AAW18670	Factor VIII-dB695-
28	127	93.4	2098	17	AA786863	Factor-VIII. Homo
29	127	93.4	2332	8	AA717126	Factor VIII:c varia
30	127	93.4	2332	8	AA717127	Factor VIII:c varia
31	127	93.4	2332	8	AA717128	Factor VIII:c varia
32	127	93.4	2332	8	AA717129	Factor VIII:c varia
33	127	93.4	2332	14	AA743257	Human Factor VIII
34	127	93.4	2332	18	AAW33222	Procoagulant-activ
35	127	93.4	2332	18	AAW33223	Procoagulant-activ
36	127	93.4	2332	18	AAW33224	Procoagulant-activ
37	127	93.4	2332	18	AAW33225	Procoagulant-activ
38	127	93.4	2332	18	AAW33226	Procoagulant-activ
39	127	93.4	2332	19	AAW53483	Human factor VIII
40	127	93.4	2332	19	AAW44132	Homo sapiens modif
41	127	93.4	2332	20	AAV31594	Human factor VIII
42	127	93.4	2332	21	AAV57847	Human Factor VIII
43	127	93.4	2332	22	AAE10826	Human mature wild-
44	127	93.4	2332	22	AAE11200	Human factor VIII
45	127	93.4	2332	22	AA711902	N-terminal truncat

ALIGNMENTS

RESULT 1	
AA73020	AA73020 standard; peptide: 368 AA.
ID	AA73020:
AC	AA73020:
DT	21-NOV-1995 (first entry)
DE	Human Factor-VIII fragment.
KW	Factor-VIII; blood-clotting; blood; coagulant; haemophilic-A.
OS	Homo sapiens.
PN	WO9513301-A.
PD	18-MAY-1995.
PF	10-NOV-1994; 94WO-DK00424.
PR	12-NOV-1993; 93DK-0001281.
PA	(NOVO) NOVO-NORDISK AS.
PI	Persson E;
DR	WPI; 1995-194038/25.
PT	Crosslinked Factor VIII polypeptide which is stable - is prep'd. using bis(sulphosuccinimidy) suberate or disuccinimidy suberate in the presence of polysorbate 80 to produce a coagulant with long lasting activity
XX	

```

PS Disclosure; Page 19; 36pp; English.
XX
CC This is a fragment corresponding to internal AAs 373-740 of human
CC Factor-VIII which may be crosslinked resulting in
CC increased stability and retention of high activity over extended
CC periods of time after activation by thrombin. The polypeptide is
CC used to prevent or treat diseases caused by the absence or deficiency
CC of Factor-VIII in a subject such as haemophilia.
XX
SQ Sequence 368 AA;

Query Match 93.4%; Score 127; DB 16; Length 368;
Best Local Similarity 96.0%; Pred. No. 2.1e-10;
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 RPNYSRRLPGVKVHLKDFPILPGEI 25
DB 112 RPLYSRRLPGVKVHLKDFPILPGEI 136

RESULT 2
AAR74088
ID AAR74088 standard; protein: 720 AA.
AC AAR74088;
XX
XX 04-NOV-1995 (first entry)
DT
XX Factor-VIII heavy chain N-terminal fragment.
DE
XX human; Factor VIII; heavy chain; N-terminal fragment;
KW thrombin cleavage; blood-clotting.
XX
XX Homo sapiens.
OS
XX WO9513300-A.
PN
XX 18-MAY-1995.
PD
XX 04-NOV-1995 (first entry)
DT
XX Factor-VIII heavy chain N-terminal fragment.
DE
XX human; Factor VIII; heavy chain; N-terminal fragment;
KW thrombin cleavage; blood-clotting.
XX
XX Homo sapiens.
OS
XX WO9513300-A.
PN
XX 18-MAY-1995.
PD
XX 10-NOV-1994; 94WO-DK00423.
PF
XX 12-NOV-1993; 93DK-0001280.
PR
XX (NOVO ) NOVO-NORDISK AS.
PA
XX Ezban Rasmussen M, Kjalke M;
PI
XX WPI; 1995-194037/25.
PS
XX
XX The sequence represents N-terminal residues 1-720 of a human Factor-
XX VIII heavy chain. The sequence is shorter than the A1-A2 domain,
XX and is produced by treating a polypeptide containing the full A1-A2
XX domain of full-length Factor-VIII with a protease, e.g. thrombin.
XX The fragment has the same specific activity as full-length
XX Factor-VIII in a chromogenic assay, but activity is a factor of two
XX lower in a clotting assay, and the fragment is activated by thrombin
XX at a slower rate and to a lower level than fragments 1-740 (AAR74090),
XX 1-729 (AAR74089) and plasma Factor-VIII. The fragment may be produced
XX recombinantly to reduce production costs and improve safety, and
XX production levels and stability are higher than for the full-length
XX form. The fragment may be used to treat patients who have developed
XX antibodies against epitopes in the C-terminal part of the heavy chain.
XX
SQ Sequence 720 AA;

Query Match 93.4%; Score 127; DB 16; Length 720;
Best Local Similarity 96.0%; Pred. No. 4.3e-10;
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 RPNYSRRLPGVKVHLKDFPILPGEI 25
DB 484 RPLYSRRLPGVKVHLKDFPILPGEI 508

RESULT 3
AAR74089
ID AAR74089 standard; protein: 729 AA.
AC AAR74089;
XX
XX 04-NOV-1995 (first entry)
DT
XX Factor-VIII heavy chain N-terminal fragment.
DE
XX human; Factor VIII; heavy chain; N-terminal fragment;
KW thrombin cleavage; blood-clotting.
XX
XX Homo sapiens.
OS
XX WO9513300-A.
PN
XX 18-MAY-1995.
PD
XX 10-NOV-1994; 94WO-DK00423.
PF
XX 12-NOV-1993; 93DK-0001280.
PR
XX (NOVO ) NOVO-NORDISK AS.
PA
XX Ezban Rasmussen M, Kjalke M;
PI
XX WPI; 1995-194037/25.
PS
XX
XX The sequence represents N-terminal residues 1-729 of a human Factor-
XX VIII heavy chain. The sequence is shorter than the A1-A2 domain,
XX and is produced by treating a polypeptide containing the full A1-A2
XX domain of full-length Factor-VIII with a protease, e.g. thrombin.
XX The fragment has the same coagulant specific activity as full-length
XX Factor-VIII in a chromogenic assay, and is activated by thrombin at
XX a similar rate. The fragment may be produced recombinantly to reduce
XX production costs and improve safety, and production levels and
XX stability are higher than for the full-length form. The fragment may
XX be used to treat patients who have developed antibodies against
XX epitopes in the C-terminal part of the heavy chain.
XX
SQ Sequence 729 AA;

Query Match 93.4%; Score 127; DB 16; Length 729;
Best Local Similarity 96.0%; Pred. No. 4.3e-10;
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 RPNYSRRLPGVKVHLKDFPILPGEI 25
DB 484 RPLYSRRLPGVKVHLKDFPILPGEI 508

RESULT 4
AAR76961
ID AAR76961 standard; protein: 740 AA.
XX
XX AAR76961;
XX

```

```

DT 09-MAR-1996 (first entry)
DE Human Factor-VIII derivative.
KW Factor-VIII; therapeutic; blood-clotting.
XX Homo sapiens.
OS
FH
XX
XX Key Location/Qualifiers
FT Misc-difference 692
FT /label= absent or Ala, Thr, Ser, Gly or Asp
FT Misc-difference 720
FT /label= absent or Gln, Ser, Thr, Val or Ala
FT Misc-difference 729
FT /label= absent or Val, Ala or Ile
FT
XX
XX WO9518827-A1.
XX
XX 13-JUL-1995.
XX
XX 06-JAN-1995; 95WO-DK00008.
XX
XX 07-JAN-1994; 94DK-0000032.
XX
XX (NOVO ) NOVO-NORDISK AS.
XX
XX Ezban Rasmussen M, Nicolaisen EM, Persson E;
XX WPI; 1995-255039/33.
XX
XX Novel factor VIII derivative used to treat haemophilia - and
XX comprises a functional A2 domain containing a mutation at one or
XX more Cys residues.
XX
XX Disclosure; Page 14-17; 30pp; English.
XX
XX The new Factor-VIII derivative comprises a functional A2 domain in
XX which Cys-692 is deleted or replaced with another amino acid
XX residue, preferably Ser (see AAR76962). Alternatively, Glu-720 and/or
XX Tyr-729 are deleted or substituted with various amino acids (as in
XX the Features). The new derivative has the same activity as the wild-
XX type Factor-VIII but with improved stability (the activity is
XX maintained for a longer period compared to the rapid decline of the
XX activity of wt Factor-VIII). The new derivative can be used in a
XX composition for treating diseases caused by an absence or deficiency
XX of Factor-VIII, especially haemophilia.
XX
XX Sequence 740 AA;
XX
XX Query Match 93.4%; Score 127; DB 16; Length 740;
XX Best Local Similarity 96.0%; Pred. No. 4.4e-10;
XX Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
XX
XX QY 1 RPNYSRRLPKGVKHLKDFPILPGEI 25
XX ||| ||||| ||||| ||||| ||||| |||||
XX Db 484 RPLYSRRLPKGVKHLKDFPILPGEI 508
XX
XX RESULT 5
XX AAR76962
XX ID AAR76962 standard; proteln; 740 AA.
XX
XX AC AAR76962;
XX
XX DT 09-MAR-1996 (first entry)
XX
XX DE Human Factor-VIII derivative.
XX
XX KW Factor-VIII; therapeutic; blood-clotting.
XX
XX OS Homo sapiens.
XX
XX FH Key Location/Qualifiers

```

```

FT Misc-difference 692
FT /label= Cys substituted by Ser
FT Misc-difference 720
FT /label= absent or Gln, Ser, Thr, Val or Ala
FT Misc-difference 729
FT /label= absent or Val, Ala or Ile
FT
XX
XX WO9518827-A1.
XX
XX 13-JUL-1995.
XX
XX 06-JAN-1995; 95WO-DK00008.
XX
XX 07-JAN-1994; 94DK-0000032.
XX
XX (NOVO ) NOVO-NORDISK AS.
XX
XX Ezban Rasmussen M, Nicolaisen EM, Persson E;
XX WPI; 1995-255039/33.
XX
XX Novel factor VIII derivative used to treat haemophilia - and
XX comprises a functional A2 domain containing a mutation at one or
XX more Cys residues.
XX
XX Disclosure; Page 18-20; 30pp; English.
XX
XX The new Factor-VIII derivative comprises a functional A2 domain in
XX which Cys-692 is replaced with Ser. For other (less preferred)
XX substitutions at this site, see AAR76961. Alternatively, Glu-720 and/or
XX Tyr-729 are deleted or substituted with various amino acids (as in
XX the Features). The new derivative has the same activity as the wild-
XX type Factor-VIII but with improved stability (the activity is
XX maintained for a longer period compared to the rapid decline of the
XX activity of wt Factor-VIII). The new derivative can be used in a
XX composition for treating diseases caused by an absence or deficiency
XX of Factor-VIII, especially haemophilia.
XX
XX Sequence 740 AA;
XX
XX Query Match 93.4%; Score 127; DB 16; Length 740;
XX Best Local Similarity 96.0%; Pred. No. 4.4e-10;
XX Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
XX
XX QY 1 RPNYSRRLPKGVKHLKDFPILPGEI 25
XX ||| ||||| ||||| ||||| ||||| |||||
XX Db 484 RPLYSRRLPKGVKHLKDFPILPGEI 508
XX
XX RESULT 6
XX AAR73021
XX ID AAR73021 standard; peptide; 740 AA.
XX
XX AC AAR73021;
XX
XX DT 21-NOV-1995 (first entry)
XX
XX DE Human Factor-VIII N-terminal fragment.
XX
XX KW Factor-VIII; blood-clotting; blood; coagulant; haemophilia-A.
XX
XX OS Homo sapiens.
XX
XX PN WO9513301-A.
XX
XX PD 18-MAY-1995.
XX
XX PF 10-NOV-1994; 94WO-DK00424.
XX
XX PR 12-NOV-1993; 93DK-0001281.
XX
XX (NOVO ) NOVO-NORDISK AS.
XX
XX

```



```

XX SQ Sequence 740 AA;
Query Match 93.4%; Score 127; DB 16; Length 740;
Best Local Similarity 96.0%; Pred. No. 4.4e-10;
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 RPNYSRRLPGVKHLKDFPILPGEI 25
Db 484 RPLYSRRLPGVKHLKDFPILPGEI 508

RESULT 9
AAW33227
ID AAW33227 standard; protein; 1383 AA.
XX AC AAW33227;
XX DT 30-APR-1998 (first entry)
XX DE Procoagulant-active human factor VIII:C (FVIII) mutant protein.
XX KW Pro-coagulant active factor VIII; FVIII; haemophilia A;
KW recombinant secretion; pro-coagulant activity; resistance;
KW activated protein C cleavage; APC; B domain; A2 domain; A3 domain;
KW von Willebrand factor binding site; blinding affinity;
KW FVIII replacement therapy.
XX OS Synthetic.
OS Homo sapiens.
XX FH Key Location/Qualifiers
FT Region 1..346 /note= "factor VIIIA heavy chain"
FT Region 741..1383 /note= "factor VIIIA light chain"
FT Domain 1..329 /note= "A1 domain"
FT Domain 1..179 /note= "A2 domain"
FT Domain 187..329 /note= "plastocyanin-like domain 1"
FT Domain /note= "plastocyanin-like domain 2"
FT Domain 380..711 /note= "A3 domain"
FT Misc_feature 711..746 /note= "a spacer of the sequence
SFSQSRHPSTKQKFNATIPENDIEKTDPMWF
AHRTPMKIQNVSSDLMMLL is inserted
between domains A2 and A3"
FT Domain 380..554 /note= "plastocyanin-like domain 3"
FT Domain 564..711 /note= "plastocyanin-like domain 4"
FT Domain 746..1073 /note= "A3 domain"
FT Domain 1073..1221 /note= "C1 domain"
FT Domain 1226..1378 /note= "C2 domain"
FT Cleavage-site 372..373 /note= "by thrombin"
FT Disulfide-bond 153..179 /note= "probable"
FT Disulfide-bond 528..554 /note= "probable"
FT Misc-difference 740 /label= R740A
FT /note= "wild type Arg replaced with Ala"
XX PN W09740145-Al.
XX PD 30-OCT-1997.
XX

```

```

PF 24-APR-1997; 97WO-US06563.
XX
XX 15-MAY-1996; 96US-0017785.
XX 24-APR-1996; 96US-0016117.
XX (UNMI ) UNIV MICHIGAN.
XX
XX Amano K, Kaufman RJ, Pipe SW;
XX WPI; 1997-535830/49.
XX
XX Modified human pro-coagulant active factor VIII - can be
XX administered to haemophiliacs, i.e. factor VIII replacement therapy
XX
XX Claim 20; Page -; 57pp; English.
XX
XX The present sequence represents a novel pro-coagulant active factor
XX VIII (FVIII) mutant protein, comprising a deletion of the B domain and
XX von Willebrand factor binding site, a mutation at Arg740 and an addition
XX of an amino acid sequence spacer between the A2 and A3 domains. Factor
XX VIII, along with calcium and phospholipid, acts as a cofactor for factor
XX IXA, when it converts factor X to the activated form (factor XA). FVIII
XX is the coagulation factor deficient in the X-chromosome-linked bleeding
XX disorder haemophilia A. Several other mutant FVIII proteins have also
XX been created (see AAW33222-29). The FVIII mutant F309S (AAW33225) is
XX capable of recombinant secretion at higher levels than typically obtained
XX with wild type FVIII and retains pro-coagulant activity. The FVIII
XX mutant R336I (AAW33222) and R562K (AAW33223) are resistant to activated
XX protein C (APC) cleavage. The present FVIII mutant can form a more
XX stable configuration, and have an approximate 5-fold increase in
XX specific activity compared to purified wild type FVIII, while increasing
XX their binding affinity to von Willebrand factor improves their
XX stability. The FVIII proteins can be administered to haemophiliacs, i.e.
XX FVIII replacement therapy, while the nucleic acid molecule can be used
XX for gene therapy.
XX note: this sequence does not appear in the specification; it was created
XX using sequences from the given references.
XX
XX SQ Sequence 1383 AA;
Query Match 93.4%; Score 127; DB 18; Length 1383;
Best Local Similarity 96.0%; Pred. No. 8.6e-10;
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 RPNYSRRLPGVKHLKDFPILPGEI 25
Db 484 RPLYSRRLPGVKHLKDFPILPGEI 508

RESULT 10
AAW33228
ID AAW33228 standard; protein; 1383 AA.
XX AC AAW33228;
XX DT 30-APR-1998 (first entry)
XX DE Procoagulant-active human factor VIII:C (FVIII) mutant protein.
XX KW Pro-coagulant active factor VIII; FVIII; haemophilia A;
KW recombinant secretion; pro-coagulant activity; resistance;
KW activated protein C cleavage; APC; B domain; A2 domain; A3 domain;
KW von Willebrand factor binding site; binding affinity;
KW FVIII replacement therapy.
XX OS Synthetic.
OS Homo sapiens.
XX FH Key Location/Qualifiers
FT Region 1..346 /note= "factor VIIIA heavy chain"
FT Region 741..1383 /note= "factor VIIIA light chain"
FT

```

FT Domain 1..329 /note= "A1 domain"
 FT Domain 1..179 /note= "plastocyanin-like domain 1"
 FT Domain 187..329 /note= "plastocyanin-like domain 2"
 FT Domain 380..711 /note= "A2 domain"
 FT Misc_feature 711..746 /note= "a spacer of the sequence
 SFQNSRHPSTROKQFNATIPENDIEKTPDPWF
 AHRTMPKIQNVSSDLMML is inserted
 between domains A2 and A3"
 FT Domain 380..554 /note= "plastocyanin-like domain 3"
 FT Domain 564..711 /note= "plastocyanin-like domain 4"
 FT Domain 746..1073 /note= "A3 domain"
 FT Domain 1073..1221 /note= "C1 domain"
 FT Domain 1226..1378 /note= "C2 domain"
 FT Cleavage-site 372..373 /note= "by thrombin"
 FT Disulfide-bond 153..179 /note= "probable"
 FT Disulfide-bond 528..554 /note= "probable"
 FT Misc-difference 336 /label= R336I /note= "wild type Arg replaced with Ile"
 FT Misc-difference 562 /label= R562K /note= "wild type Arg replaced with Lys"
 FT Misc-difference 740 /label= R740A /note= "wild type Arg replaced with Ala"
 FT XX W09740145-A1.
 FT XX 30-OCT-1997.
 FT XX 24-APR-1997; 97WO-US06563.
 FT XX 15-MAY-1996; 96US-0017785.
 FT XX 24-APR-1996; 96US-0016117.
 FT XX (UNMI) UNIV MICHIGAN.
 FT XX Amano K, Kaufman RJ, Pipe SW;
 FT WP1; 1997-535830/49.
 FT Modified human pro-coagulant active factor VIII - can be
 FT administered to haemophiliacs, i.e. factor VIII replacement therapy
 FT Claim 18; Page -: 57pp; English.
 FT The present sequence represents a novel pro-coagulant active factor
 FT VIII (FVIII) mutant protein, comprising a deletion of the B domain and
 FT von Willebrand factor binding site, mutations R336I, R562K and R740A and
 FT an addition of an amino acid sequence spacer between the A2 and A3
 FT domains. Factor VIII, along with calcium and phospholipid, acts as a
 FT cofactor for factor IXa, when it converts factor X to the activated form
 FT (factor Xa). FVIII is the coagulation factor deficient in the
 FT X-chromosome-linked bleeding disorder haemophilia A. Several other
 FT mutant FVIII proteins have also been created (see AAW33222-29). The FVIII
 FT mutant F309S (AAW33225) is capable of recombinant secretion at higher
 FT levels than typically obtained with wild type FVIII and retains
 FT pro-coagulant activity. The FVIII mutant R336I (AAW33222) and R562K
 FT (AAW33223) are resistant to activated protein C (APC) cleavage. The FVIII
 FT mutant comprising a deletion of the B domain and von Willebrand factor

CC binding site, a mutation at Arg740 and an addition of an amino acid
 CC sequence spacer between the A2 and A3 domains can form a more stable
 CC configuration, and have an approximate 5-fold increase in specific
 CC activity compared to purified wild type FVIII, while increasing their
 CC binding affinity to von Willebrand factor improves their stability.
 CC The FVIII proteins can be administered to haemophiliacs, i.e. FVIII
 CC replacement therapy, while the nucleic acid molecule can be used for
 CC gene therapy.
 CC note: this sequence does not appear in the specification; it was created
 CC using sequences from the given references.
 XX XX
 SQ Sequence 1383 AA;
 Query Match 93.4%; Score 127; DB 18; Length 1383;
 Best Local Similarity 96.0%; Pred No. 8.6e-10;
 Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 RPNYSRRLLPKGVKHLKDFPILPGEI 25
 Db 484 RPLYSRRLLPKGVKHLKDFPILPGEI 508
 || |||||
 RESULT 11
 AAW33229
 ID AAW33229 standard; protein: 1383 AA.
 XX AC AAW33229;
 XX 30-APR-1998 (first entry)
 DE Procoagulant-active human factor VIII:C (FVIII) mutant protein.
 XX KW recombinant active factor VIII: FVIII: haemophilia A;
 KW recombinant secretion; pro-coagulant activity; resistance;
 KW activated protein C cleavage; APC; B domain; A2 domain; A3 domain;
 KW von Willebrand factor binding site; binding affinity;
 KW FVIII replacement therapy.
 XX OS Synthetic.
 OS Homo sapiens.
 XX XX
 FH Key Location/Qualifiers
 FT Region 1..346
 FT /note= "factor VIIIA heavy chain"
 FT Region 741..1383
 FT /note= "factor VIIIA light chain"
 FT Domain 1..329
 FT /note= "A1 domain"
 FT Domain 1..179
 FT /note= "plastocyanin-like domain 1"
 FT Domain 187..329
 FT /note= "plastocyanin-like domain 2"
 FT Domain 380..711
 FT /note= "A2 domain"
 FT Misc_feature 711..746
 FT /note= "a spacer of the sequence
 SFQNSRHPSTROKQFNATIPENDIEKTPDPWF
 AHRTMPKIQNVSSDLMML is inserted
 between domains A2 and A3"
 FT Domain 380..554
 FT /note= "plastocyanin-like domain 3"
 FT Domain 564..711
 FT /note= "plastocyanin-like domain 4"
 FT Domain 746..1073
 FT /note= "A3 domain"
 FT Domain 1073..1221
 FT /note= "C1 domain"
 FT Domain 1226..1378
 FT /note= "C2 domain"
 FT Cleavage-site 372..373
 FT /note= "by thrombin"
 FT Disulfide-bond 153..179
 FT /note= "probable"

FT Disulfide-bond 528..554 /note= "probable"
 FT Misc-difference 309 /label= F309S
 FT /note= "wild type Phe replaced with Ser"
 FT Misc-difference 740
 FT /label= R740A
 FT /note= "wild type Arg replaced with Ala"
 XX
 XX WO9740145-A1.
 XX
 XX 30-OCT-1997.
 XX
 XX 24-APR-1997; 97WO-US06563.
 XX
 XX 15-MAY-1996; 96US-0017785.
 XX 24-APR-1996; 96US-0016117.
 XX
 XX (UNMI) UNIV MICHIGAN.
 XX
 XX Amano K, Kaufman RJ, Pipe SW;
 XX WPI; 1997-535830/49.
 XX
 XX Modified human pro-coagulant active factor VIII - can be
 XX administered to haemophiliacs, i.e. factor VIII replacement therapy
 XX
 XX Claim 19; Page -: 57pp; English.
 XX
 XX The present sequence represents a novel pro-coagulant active factor
 XX FVIII (FVIII) mutant protein, comprising a deletion of the B domain and
 XX von Willebrand factor binding site, mutations F309S, R740A and addition
 XX of an amino acid sequence spacer between the A2 and A3 domains. Factor
 XX FVIII, along with calcium and phospholipid, acts as a cofactor for factor
 XX IXA, when it converts factor X to the activated form (factor XA). FVIII
 XX is the coagulation factor deficient in the X-chromosome-linked bleeding
 XX disorder haemophilia A. Several other mutant FVIII proteins have also
 XX been created (see AAW33222-29). The FVIII mutant F309S (AAW33225) is
 XX capable of recombinant secretion at higher levels than typically obtained
 XX with wild type FVIII and retains pro-coagulant activity. The FVIII
 XX mutant R336I (AAW33222) and R562K (AAW33223) are resistant to activated
 XX protein C (APC) cleavage. The present FVIII mutant can form a more
 XX stable configuration, and have an approximate 5-fold increase in
 XX specific activity compared to purified wild type FVIII, while increasing
 XX their binding affinity to von Willebrand factor improves their
 XX stability. The FVIII proteins can be administered to haemophiliacs, i.e.
 XX FVIII replacement therapy, while the nucleic acid molecule can be used
 XX for gene therapy.
 XX note: this sequence does not appear in the specification; it was created
 XX using sequences from the given references.
 XX
 XX SO Sequence 1383 AA;
 Query Match 93.4%; Score 127; DB 18; Length 1383;
 Best Local Similarity 96.0%; Pred. No. 8.6e-10;
 Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 RPNYSRRLPGKGVKHLKDFPILPGEI 25
 DB 484 RPLYSRRLPGKGVKHLKDFPILPGEI 508
 RESULT 12
 AAP80268
 ID AAP80268 standard; protein; 1424 AA.
 XX
 XX AAP80268;
 XX
 XX 10-OCT-1990 (first entry)
 XX
 XX Modified factor VIII:C sequence with the R740-D1658 deletion.
 XX
 XX Modified factor VIII:C; maturation polypeptide; haemophilia;

KW blood coagulation; RD deletion; procoagulant.
 XX Homo sapiens.
 XX PN WO8800831-A.
 XX PD 11-FEB-1988.
 XX PF 31-JUL-1987; 87WO-US01814.
 XX PR 01-AUG-1986; 86US-0893375.
 XX PA (BIOJ) BIOGEN NV (PASE/).
 XX PI Pasek MP;
 XX WPI; 1988-049866/07.
 XX N-PSDB; AAN80447.
 XX
 XX New DNA sequences encoding modified factor VIII:C - with deletion of DNA
 XX encoding maturation polypeptide, useful for high yield transformation.
 XX
 XX Claim 3; Page 60-61-62-63; 97pp; English.
 XX
 XX The RD deletion removes the DNA from Ser 741 to Ser 1657.
 XX A major part of the sequence encoding the maturation polypeptide of
 XX factor VIII:C is deleted, i.e. Gln 744 - Asp 1563.
 XX The full length Factor VIII:C cDNA has two changes with respect to the
 XX published sequence (EPO application 160457):
 XX CTC to CTA at Leu 242 and TTC to CTC change at amino acid residue 1880
 XX (Phe to Leu). The product is produced in approx. 20 times higher
 XX yields than previous recombinant produced factor VIII:C and are more
 XX easily purified. The peptide is used for treating haemophilia A, both
 XX acute and prolonged bleeding.
 XX See also AAN80444 and AAN80446.
 XX
 XX SO Sequence 1424 AA;
 Query Match 93.4%; Score 127; DB 9; Length 1424;
 Best Local Similarity 96.0%; Pred. No. 8.8e-10;
 Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 RPNYSRRLPGKGVKHLKDFPILPGEI 25
 DB 484 RPLYSRRLPGKGVKHLKDFPILPGEI 508
 RESULT 13
 AAP91169
 ID AAP91169 standard; protein; 1424 AA.
 XX
 XX AAP91169;
 XX
 XX 26-JUN-1990 (first entry)
 XX
 XX Sequence of 740 Arg-1649 Glu human Factor VIII:C.
 XX
 XX Human Factor VIII:C; Ad.RE.neo; 740 Arg-1649 Glu Factor VIII:C;
 XX haemophilia A.
 XX
 XX Homo sapiens.
 XX EP306968-A.
 XX PD 15-MAR-1989.
 XX PF 09-SEP-1988; 88EP-0114769.
 XX PR 08-APR-1988; 88JP-0085454.
 XX
 XX (KAGA) CEMO-SERO-THERAP (TEIJ).
 XX
 XX Sugiyama T, Masuda K, Tajima Y, Yonemura H;

XX WPI: 1989-078467/11.
 DR N-PSDB; AAN90654.
 XX
 PT Prodn. of recombinant human Factor-VIII-C -
 PT using animal cells transformed with a vector contg. the gene for
 PT Factor VIII:C and a promoter
 XX
 PS Fig 1(1) - 1(13); : 32pp; English.
 XX
 CC Arg-740 of the carboxyl terminus of the H chain is
 CC directly bonded by a peptide bond to Glu-1649 of the amino terminus of
 CC L chain. A pred. expression vector used to transform animal cell so
 CC that they produce human Factor VIII:Cis plasmid Ad.RE.neo.
 CC The expression vector has at least one promoter upstream of AAN90654.
 CC The transformants can constantly and continuously produce human Factor
 CC VIII:C in high yield on a commercial scale. The human Factor VIII:C so
 CC produced is considered to corresp. to the smallest species of active and
 CC intact Factor VIII:C molecules in the human blood plasma. It is useful
 CC for treating haemophilia A patients.
 XX
 SQ Sequence 1424 AA;
 Query Match 93.4%; Score 127; DB 10; Length 1424;
 Best Local Similarity 96.0%; Pred. No. 8.8e-10;
 Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 Qy 1 RPNYSRRLPGVKVHLKDFPILPGEI 25
 Db 484 RPLYSRRLPGVKVHLKDFPILPGEI 508
 RESULT 14
 AAB48842
 ID AAB48842 standard; protein; 1424 AA.
 XX
 AC AAB48842;
 XX
 DT 13-MAR-2001 (first entry)
 XX
 DE Mutant mature human factor VIII, SEQ ID NO:5.
 XX
 KW Factor VIII; human; B domain; LRP-mediated plasma clearance;
 KW receptor-dependent clearance; receptor-independent clearance;
 KW half-life; haemophilia; mutant; mutuin.
 XX
 OS Homo sapiens.
 XX
 PN WO200071714-A2.
 XX
 PD 30-NOV-2000.
 XX
 PF 24-MAY-2000; 2000WO-US14111.
 XX
 PR 24-MAY-1999; 99US-0135847.
 XX
 PA (AMNA-) AMERICAN NAT RED CROSS.
 XX
 PT Saenko EL, Strickland DK;
 XX
 PS WPI: 2001-025163/03.
 XX
 CC Factor VIII mutants having increased half-life useful for treating
 CC hemophilia, comprise one or more amino acid substitutions in the A2
 CC and/or C2 domain of factor VIII -
 XX
 PS Claim 9; Fig 2A-B; 121pp; English.
 XX
 CC The invention relates to human factor VIII mutants comprising an amino
 CC acid substitution at one or more positions in the A2 domain and/or an
 CC amino acid substitution at one or more positions in the C2 domain.
 CC The invention also encompasses a factor VIII mutant which lacks a B
 CC domain (AAB48842). The factor VIII mutants have an increased half-life

CC in the bloodstream. The A2 domain mutants exhibit reduced LRP-dependent
 CC (receptor-dependent) clearance of factor VIII, while C2 domain mutants
 CC have reduced receptor-independent clearance. The invention also relates
 CC to a method of using RAP (receptor associated protein), a protein which
 CC inhibits LRP (low density lipoprotein related protein)-mediated ligand
 CC internalisation, to increase the half-life of factor VIII. The mutant
 CC factor VIII proteins, and nucleotides encoding them, are useful
 CC for treating haemophilia. RAP, LRP-binding RAP mutants or fragments, and
 CC nucleic acids encoding them may also be used in the treatment of
 CC haemophilia. In combination with a mutant factor VIII protein or DNA of
 CC the invention. The invention provides means of increasing the half-life
 CC of factor VIII by reducing its clearance from plasma. The present
 CC sequence represents a mutant mature human factor VIII which lacks a B
 CC domain.
 XX
 SQ Sequence 1424 AA;
 Query Match 93.4%; Score 127; DB 22; Length 1424;
 Best Local Similarity 96.0%; Pred. No. 8.8e-10;
 Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 Qy 1 RPNYSRRLPGVKVHLKDFPILPGEI 25
 Db 484 RPLYSRRLPGVKVHLKDFPILPGEI 508
 RESULT 15
 AAP80267
 ID AAP80267 standard; protein; 1425 AA.
 XX
 AC AAP80267;
 XX
 DT 10-OCT-1990 (first entry)
 XX
 DE Modified factor VIII:C sequence with the R740-E1649 deletion.
 XX
 KW Modified factor VIII:C; maturation polypeptide; haemophilia;
 KW blood coagulation; RE deletion.
 XX
 OS Homo sapiens.
 XX
 PN WO8800831-A.
 XX
 PD 11-FEB-1988.
 XX
 PF 31-JUL-1987; 87WO-US01814.
 XX
 PR 01-AUG-1986; 86US-0893375.
 XX
 PA (BIOJ) BIOGEN NV (PASE/).
 XX
 PT Pasek MP;
 XX
 PS WPI: 1988-049866/07.
 XX
 N-PSDB; AAN80446.
 XX
 CC New DNA sequences encoding modified factor VIII:C - with deletion of DNA
 CC encoding maturation polypeptide, useful for high yield transformation.
 XX
 PS Claim 3; Page 57-58-59-60; 97pp; English.
 XX
 CC The entire sequence encoding the maturation polypeptide of
 CC factor VIII:C is deleted, i.e. Arg 740-Glu 1649.
 CC The full length Factor VIII:C cDNA has two changes with respect to the
 CC published sequence (EPO application 160457):
 CC CIG to CTA at Leu 242 and TTC to CTC change at amino acid residue 1880
 CC (Phe to Leu). The product is produced in approx. 20 times higher
 CC yields than previous recombinant produced factor VIII:C and are more
 CC easily purified. The peptide is used for treating haemophilia A, both
 CC acute and prolonged bleeding.
 CC See also AAN80444 and AAN80447.
 XX
 PS Sequence 1425 AA;

Query Match 93.4%; Score 127; DB 9; Length 1425;
Best Local Similarity 96.0%; Pred. No. 8.8e-10;
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 RPNYSRRLPKGVKHLKDFPILPGEI 25
|| |||||
Db 485 RPLYSRRLPKGVKHLKDFPILPGEI 509

Search completed: January 23, 2003, 09:00:54
Job time : 38 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 23, 2003, 09:01:18 ; Search time 55 seconds
(without alignments)
43.697 Million cell updates/sec

Title: SEQ2-ASNAT3

Perfect score: 136

Sequence: 1 RPNYSRRLLPKGVKHLKDFILPGEI 25

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR_73:*

1: pir1:*

2: pir2:*

3: pir3:*

4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	127	93.4	2351	1 EZHU	coagulation factor
2	94	69.1	2319	2 A47004	coagulation factor
3	75	55.1	2133	2 T42763	coagulation factor
4	51	37.5	1607	2 T04583	TMV resistance pro
5	50	36.8	958	2 A82583	conserved hypothet
6	49	36.0	412	2 G71070	hypothetical prote
7	48.5	35.7	400	2 JC4591	alpha-1,3-fucosyl
8	48.5	35.7	405	2 B36340	alpha-1,3-fucosyl
9	48.5	35.7	433	2 A57596	alpha-1,3-fucosyl
10	48.5	35.7	1253	2 T45787	disease resistance
11	48	35.3	156	2 B83164	conserved hypothet
12	48	35.3	264	2 AC2515	hypothetical prote
13	48	35.3	600	2 B46642	DNA-directed DNA p
14	48	35.3	1690	2 T40847	probable RNA biog
15	47	34.6	444	2 S54011	1-aminocyclopropan
16	46	33.8	308	2 S67657	hypothetical prote
17	46	33.8	348	2 G71681	hypothetical prote
18	46	33.8	467	2 T38353	serine hydroxymeth
19	46	33.8	480	2 A99164	hypothetical prote
20	46	33.8	519	2 S75752	oligopeptide trans
21	46	33.8	885	2 T38387	hypothetical prote
22	46	33.8	1001	2 C88779	protein T2003.9 [1
23	46	33.8	1023	2 A59431	KIAA0013 protein [
24	46	33.8	1038	2 T25033	hypothetical prote
25	45.5	33.5	161	2 F75633	hypothetical prote
26	45.5	33.5	246	2 H82553	3-demethylubiquino
27	45.5	33.5	554	2 B80679	3-(3-hydroxyphenyl
28	45.5	33.5	554	2 F85529	3-(3-hydroxyphenyl
29	45.5	33.5	554	2 C64762	probable monooxyge

RESULT 1

EZHU

coagulation factor VIII precursor [validated] - human

N:Alternate names: antihemophilic factor A; coagulation factor VIIIc; procoagulant co
C:Species: Homo sapiens (man)
C:Date: 28-Aug-1985 #sequence_revision 28-Aug-1985 #text_change 08-Dec-2000
C:Accession: I54318; A00525; I58059; A23584; A26174; A42348; A43986; S63527; S66445;
R:Gitschier, J.; Wood, W.I.
Hum. Mol. Genet. 1, 199-200, 1992

A:Title: Sequence of the exon-containing regions of the human factor VIII gene.

A:Reference number: I54318; MUID:93265012; PMID:1303178

A:Accession: I54318

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-1921, S',1923-2351 <RES>

A:Cross-references: GB:M88648; NID:g182381; PIDN:AAA52420.1; PID:g182383

R:Wood, W.I.; Capon, D.J.; Simonsen, C.C.; Eaton, D.L.; Gitschier, J.; Keyt, B.; Seeb
Nature 312, 330-337, 1984

A:Title: Expression of active human factor VIII from recombinant DNA clones.

A:Reference number: A00525; MUID:85061548; PMID:6438526

A:Accession: A00525

A:Molecule type: mRNA

A:Residues: 1-2351 <WOO>

A:Cross-references: EMBL:X01165; EMBL:X01166; EMBL:X01179

R:Toole, J.J.; Koopf, J.L.; Wozney, J.M.; Sultzman, L.A.; Buecker, J.L.; Pittman, D.D
S, D.N.; Hewick, R.M.
Nature 312, 342-347, 1984

A:Title: Molecular cloning of a cDNA encoding human antihemophilic factor.

A:Reference number: I58059; MUID:85061550; PMID:6438528

A:Accession: I58059

A:Status: nucleic acid sequence not shown; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-74, V',76-1259, E',1261-2351 <RE2>

A:Cross-references: GB:X01740; NID:g182802; PIDN:AAA52484.1; PID:g182803

R:Truett, M.A.; Blacher, R.; Burke, R.L.; Caput, D.; Chu, C.; Dina, D.; Hartog, K.; K
.B.; Randolph, A.; Urdea, M.S.; Valenzuela, P.; Dahl, H.H.; Favalaro, J.; Hansen, J.;
DNA 4, 333-349, 1985

A:Title: Characterization of the polypeptide composition of human factor VIII:C and t

A:Reference number: A23584; MUID:86081184; PMID:3935400

A:Accession: A23584

A:Molecule type: mRNA

A:Residues: 1-2351 <TRU>

A:Cross-references: GB:M14113; NID:g182817; PIDN:AAA52485.1; PID:g182818

R:Eaton, D.; Rodriguez, H.; Vehar, G.A.

Biochemistry 25, 505-512, 1986

A:Title: Proteolytic processing of human factor VIII. Correlation of specific cleavag

ity.

A:Reference number: A26174; MUID:86159740; PMID:3082357

A:Accession: A26174

A:Molecule type: protein

A:Residues: 20-36;392-399, X',401-402;1668-1678;1709-1722, 'D',1723-1725;1741-1755 <EA
R:Pittman, D.D.; Wang, J.H.; Kaufman, R.J.

ALIGNMENTS

[illegible]


```

Db 314 RANERFVPRGAFIHVDFFP 333
|||||:|:|:|:|:|:|

RESULT 8
B36340
Alpha(1,3)-fucosyltransferase (EC 2.4.1.-) 4 precursor [validated] - human
N:Alternate names: CD15; ELAM-1 ligand fucosyltransferase (ELFT); FCT3A; FUC-TIV; myeloid
C:Species: Homo sapiens (man)
C:Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 20-Apr-2000
C:Accession: B36340; A36340; A41202
R:Goelz, S.E.; Hession, C.; Goff, D.; Griffiths, B.; Tizard, R.; Newman, B.; Chi-Rosso;
Cell 63, 1349-1356, 1990
A:Title: ELFT: A gene that directs the expression of an ELAM-1 ligand.
A:Reference number: A36340; MUID:91084863; PMID:1702034
A:Accession: B36340
A:Molecule type: mRNA
A:Residues: 1-405 <GOE1>
A:Cross-references: GB:M58596; NID:g182068; PIDN:AAA63172.1; PID:g182069
A:Accession: A36340
A:Molecule type: mRNA
A:Residues: 'MRLMGAAKPSGAGWEKEMAEAPQEPAGWSGRGLGPR','SGRKGRVPGWAGWPAHLAARPARHLGGAGC
A:Cross-references: GB:M58597; NID:g182070; PIDN:AAA63173.1; PID:g182071
A:Note: The codon used as an initiator for this translation is not in a good context for
R:Low, J.B.: Kukowska-Latallo, J.F.; Nair, R.P.; Larsen, R.D.; Marks, R.M.; Macher, B.A
J. Biol. Chem. 266, 17467-17477, 1991
A:Title: Molecular cloning of a human fucosyltransferase gene that determines expression
A:Reference number: A40976; MUID:91373370; PMID:1716630
A:Accession: A40976
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-86,'p',88-405 <LOW>
A:Cross-references: GB:M65030; NID:g182791; PIDN:AAA92977.1; PID:g1236720
R:Kumar, R.; Potvin, B.; Muller, W.A.; Stanley, P.
J. Biol. Chem. 266, 21777-21783, 1991
A:Title: Cloning of a human alpha(1,3)-fucosyltransferase gene that encodes ELFT but doe
A:Reference number: A41202; MUID:92042084; PMID:1718983
A:Accession: A41202
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-240,'D',242-400 <KUM>
A:Cross-references: GB:S65161; NID:g239005; PIDN:AAB20349.1; PID:g239006
C:Genetics:
A:Gene: GDB:FUT4; CD15; FCT3A; FUC-TIV
A:Cross-references: GDB:131563; OMIM:104230
A:Map position: 11q21-11q21
C:Superfamily: galactoside 3(4)-L-fucosyltransferase
C:Keywords: glycoprotein; glycosyltransferase; hexosyltransferase
F:1-48/Domain: signal sequence #status predicted <SIG>
F:49-405/Product: alpha(1,3)-fucosyltransferase 4 #status predicted <MAT>
F:91,190/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 35.7%; Score 48.5; DB 2; Length 405;
Best Local Similarity 50.0%; Pred. No. 23;
Matches 10; Conservative 3; Mismatches 6; Indels 1; Gaps 1;

QY 1 RPNYSRRLPKGV-KHLKDFP 19
|||||:|:|:|:|:|:|

Db 319 RANERFVPRGAFIHVDFFP 338
|||||:|:|:|:|:|:|

RESULT 9
A57596
Alpha-1,3-fucosyltransferase Fuc-TIV (EC 2.4.1.-) - mouse
N:Alternate names: ELAM-1 ligand fucosyltransferase homolog
C:Species: Mus musculus (house mouse)
C:Date: 08-Feb-1996 #sequence_revision 08-Feb-1996 #text_change 11-Jan-2000
C:Accession: A57596
R:Gersten, K.M.; Natsuka, S.; Trinchera, M.; Petryniak, B.; Kelly, R.J.; Hiraiwa, N.; Je
J. Biol. Chem. 270, 25047-25056, 1995
A:Title: Molecular cloning, expression, and tissue-specific exp
rase.
A:Reference number: A57596; MUID:96027607; PMID:7559635

```

```

A:Accession: A57596
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-433 <GER>
A:Cross-references: GB:U33457; NID:g1039426; PIDN:AAC52269.1; PID:g1039427
C:Superfamily: galactoside 3(4)-L-fucosyltransferase
C:Keywords: glycosyltransferase; hexosyltransferase

Query Match 35.7%; Score 48.5; DB 2; Length 433;
Best Local Similarity 50.0%; Pred. No. 24;
Matches 10; Conservative 3; Mismatches 6; Indels 1; Gaps 1;

QY 1 RPNYSRRLPKGV-KHLKDFP 19
|||||:|:|:|:|:|:|

Db 347 RANERFVPRGAFIHVDFFP 366
|||||:|:|:|:|:|:|

RESULT 10
T45787
disease resistance-like protein - Arabidopsis thaliana
N:Alternate names: protein F26013.200
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 04-Feb-2000 #sequence_revision 04-Feb-2000 #text_change 04-Feb-2000
C:Accession: T45787
R:Delveny, M.; Berger, C.; Cooke, R.; Grellet, F.; Laudie, M.; Mewes, H.W.; Lemcke, K
submitted to the Protein Sequence Database, December 1999
A:Reference number: Z23013
A:Accession: T45787
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1253 <DEL>
A:Cross-references: EMBL:AL133452
A:Experimental source: cultivar Columbia; BAC clone F26013
C:Genetics:
A:Map position: 3
A:Introns: 138/2; 490/3; 589/3; 1047/3
A:Note: F26013.200

Query Match 35.7%; Score 48.5; DB 2; Length 1253;
Best Local Similarity 39.3%; Pred. No. 76;
Matches 11; Conservative 4; Mismatches 9; Indels 9; Gaps 1;

QY 3 NYSRRLPKGVK-----HUKDFPIL 21
||:|||||:|:|:|:|:|:|

Db 533 HVALHLPKGVKSLPEELRLHWHFPLL 560
||:|||||:|:|:|:|:|:|

RESULT 11
B83164
conserved hypothetical protein PA3847 [imported] - Pseudomonas aeruginosa (strain PAO
C:Species: Pseudomonas aeruginosa
C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C:Accession: B83164
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.;
Adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; L
; Loy, S.; Olson, M.V.
Nature 406, 959-964, 2000
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pa
A:Reference number: AB2950; MUID:20437337; PMID:10984043
A:Accession: B83164
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-156 <STO>
A:Cross-references: GB:AE004802; GB:AE004091; NID:99950021; PIDN:AAG07234.1; GSPDB:GN
A:Experimental source: strain PA01
C:Genetics:
A:Gene: PA3847

Query Match 35.3%; Score 48; DB 2; Length 156;
Best Local Similarity 36.4%; Pred. No. 9.7;
Matches 8; Conservative 6; Mismatches 8; Indels 0; Gaps 0;

QY 1 RPNYSRRLPKGVKHLKDFPILP 22

```


GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 23, 2003, 08:55:57 ; Search time 11 Seconds
(without alignments)
94.264 Million cell updates/sec

Title: SEQ2-ASNAT3
Perfect score: 136
Sequence: 1 RPNYSRRLPGVKHLKDFPILPGEI_25

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	127	93.4	2351	1	FA8_HUMAN
2	94	69.1	2319	1	FA8_MOUSE
3	75	55.1	2133	1	FA8_PIG
4	48.5	35.7	405	1	FUT4_HUMAN
5	48.5	35.7	433	1	FUT4_MOUSE
6	48	35.3	495	1	DPO2_RAT
7	48	35.3	600	1	DPO2_MOUSE
8	46	33.8	467	1	GLYD_SCHPO
9	45.5	33.5	246	1	UBIG_XYLFA
10	45.5	33.5	554	1	MHPA_ECOLI
11	45	33.1	438	1	MURA_SYNY3
12	45	33.1	1032	1	ITA4_XENLA
13	45	33.1	1678	1	CA64_HUMAN
14	45	33.1	1690	1	CA44_HUMAN
15	45	33.1	2054	1	YCF2_PINTH
16	44.5	32.7	788	1	SYFB_THEMEA
17	44	32.4	130	1	RR11_MARPO
18	44	32.4	300	1	ERA_MYCTU
19	44	32.4	356	1	CDA_CAEEL
20	44	32.4	385	1	VATC_METTH
21	44	32.4	1143	1	CALI_HUMAN
22	43.5	32.0	251	1	RPC8_YEAST
23	43.5	32.0	352	1	AROB_SULTO
24	43.5	32.0	649	1	VATI_CHLTR
25	43	31.6	146	1	Y237_AQUAE
26	43	31.6	156	1	RS10_LUMRU
27	43	31.6	165	1	RS10_XENLA
28	43	31.6	166	1	RS10_ICTEPU
29	43	31.6	227	1	FLPA_PYRHO
30	43	31.6	441	1	CYSP_THEAN
31	43	31.6	529	1	TIM_SALPO
32	43	31.6	529	1	TIM_SALTY
33	43	31.6	579	1	ILVB_SPIPL

34	43	31.6	891	1	SVV_PYRAB	Q9uy55 pyrococcus
35	43	1664	1	RPA1_YEAST	P10964 saccharomyc	
36	43	31.6	3866	1	HRX_MOUSE	P55200 mus musculus
37	42.5	31.2	145	1	RS15_XIPMA	P70066 xiphophorus
38	42.5	31.2	188	1	CR12_HUMAN	P51864 homo sapien
39	42.5	31.2	649	1	VATI_CHLMU	Q9pk88 chlamydia m
40	42.5	31.2	1342	1	RPOB_ECOLI	P00575 escherichia
41	42.5	31.2	1342	1	RPOB_SALTY	P06173 salmonella
42	42	30.9	138	1	RR11_TOBAC	P06365 nicotiana t
43	42	30.9	209	1	ENGB_NEIGO	O05132 neisseria g
44	42	30.9	209	1	ENGB_NEIMA	Q93jw1 neisseria m
45	42	30.9	321	1	PLSX_CHLTR	O84817 chlamydia t

ALIGNMENTS

RESULT 1
FA8_HUMAN STANDARD; PRT: 2351 AA.
AC P00451;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Coagulation factor VIII precursor (procoagulant component)
DE (Antihemophilic factor) (AHF).
GN F8 OR F8C.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=86081164; PubMed=3935400;
RA Truett M.A., Blacher R., Burke R.L., Caput D., Chu C., Dina D.,
RA Harthog K., Kuo C.H., Maslarz F.R., Merryweather J.P., Najarian R.,
RA Pachl C., Potter S.J., Puma J., Quiroga M., Rall L.B., Randolph A.,
RA Urdea M.S., Valenzuela P., Dahl H.-H.M., Favalaro J., Hansen J.,
RA Nordfang O., Ezban M.;
RT "Characterization of the polypeptide composition of human factor
VIII:C and the nucleotide sequence and expression of the human kidney
cDNA.";
RL DNA 4:333-349(1985).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=85061548; PubMed=6438526;
RA Wood W.I., Capon D.J., Simonsen C.C., Eaton D.L., Gitschier J.,
RA Keyt B., Seeburg P.H., Smith D.H., Hollingshead P., Wion K.L.,
RA Delwart E., Tuddenham E.G.D., Vehar G.A., Lawn R.M.;
RT "Expression of active human factor VIII from recombinant DNA clones.";
RL Nature 312:330-337(1984).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=85061550; PubMed=6438528;
RA Toole J.J., Knopf J.L., Wozney J.M., Sultzman L.A., Buecker J.L.,
RA Pittman D.D., Kaufman R.J., Brown E., Shoemaker C., Orr E.C.,
RA Amphlett G.W., Foster W.B., Coe M.L., Knutson G.J., Fass D.N.,
RA Hewick R.M.;
RT "Molecular cloning of a cDNA encoding human antihemophilic factor.";
RL Nature 312:342-347(1984).
RN [4]
RP SEQUENCE FROM N.A.
RX MEDLINE=93265012; PubMed=1303178;
RA Gitschier J., Wood W.I.;
RT "Sequence of the exon-containing regions of the human factor VIII
gene.";
RL Hum. Mol. Genet. 1:199-200(1992).
RN [5]
RP SEQUENCE OF 2064-2070 FROM N.A.
RA de Water N.S., Williams R., Browett P.J.;
RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.
RN [6]
RP SULFATION OF TYR-1699.

RX MEDLINE-91093266; PubMed-1898735;
 RA Leyte A., van Schijndel H.B., Niehrs C., Huttner W.B., Verbeest M.P.,
 RA Mertens K., van Mourik J.A.;
 RT "Sulfation of Tyr1680 of human blood coagulation factor VIII is
 RT essential for the interaction of factor VIII with von Willebrand
 RT factor.";
 RL J. Biol. Chem. 266:740-746(1991).
 RN [7]
 RP SULFATION.
 RX MEDLINE-92207952; PubMed-1554716;
 RA Pittman D.D., Wang J.H., Kaufman R.J.;
 RT "Identification and functional importance of tyrosine sulfate
 RT residues within recombinant factor VIII.";
 RL Biochemistry 31:3315-3323(1992).
 RN [8]
 RP STRUCTURE BY NMR OF 2322-2343.
 RX MEDLINE-95200924; PubMed-7893714;
 RA Gilbert G.E., Baleja J.D.;
 RT "Membrane-binding peptide from the C2 domain of factor VIII forms an
 RT amphipathic structure as determined by NMR spectroscopy.";
 RL Biochemistry 34:3022-3031(1995).
 RN [9]
 RP REVIEW ON MOLECULAR BASIS OF HEMOPHILIA A.
 RX MEDLINE-91221499; PubMed-1902642;
 RA Gitschier J.;
 RT "The molecular basis of hemophilia A.";
 RL Ann. N.Y. Acad. Sci. 614:89-96(1991).
 RN [10]
 RP REVIEW ON MOLECULAR BASIS OF HEMOPHILIA A.
 RX MEDLINE-89088506; PubMed-2491949;
 RA White G.C. II, Shoemaker C.B.;
 RT "Factor VIII gene and hemophilia A.";
 RL Blood 73:1-12(1989).
 RN [11]
 RP REVIEW ON MOLECULAR BASIS OF HEMOPHILIA A.
 RX MEDLINE-95245332; PubMed-7728145;
 RA Antonarakis S.E., Kazazian H.H., Tuddenham E.G.D.;
 RT "Molecular etiology of factor VIII deficiency in hemophilia A.";
 RL Hum. Mutat. 5:1-22(1995).
 RN [12]
 RP VARIANT GLN-2326.
 RX MEDLINE-86235434; PubMed-3012775;
 RA Gitschier J., Wood W.I., Shuman M.A., Lawn R.M.;
 RT "Identification of a missense mutation in the factor VIII gene of a
 RT mild hemophiliac.";
 RL Science 232:1415-1416(1986).
 RN [13]
 RP VARIANT PRO-2135.
 RX MEDLINE-88096539; PubMed-3122181;
 RA Levinson B., Janco R.L., Phillips J.A. III, Gitschier J.;
 RT "A novel missense mutation in the factor VIII gene identified by
 RT analysis of amplified hemophilia DNA sequences.";
 RL Nucleic Acids Res. 15:9797-9805(1987).
 RN [14]
 RP VARIANT GLN-2228.
 RX MEDLINE-88191889; PubMed-2833855;
 RA Youssoufian H., Antonarakis S.E., Bell W., Griffin A.M.,
 RA Kazazian H.H.;
 RT "Nonsense and missense mutations in hemophilia A: estimate of the
 RT relative mutation rate at CG dinucleotides.";
 RL Am. J. Hum. Genet. 42:718-725(1988).
 RN [15]
 RP VARIANT GLY-291.
 RX MEDLINE-88220354; PubMed-2835904;
 RA Youssoufian H., Wong C., Aronis S., Platokoukis H., Kazazian H.H. Jr.,
 RA Antonarakis S.E.;
 RT "Moderately severe hemophilia A resulting from Glu-->Gly substitution
 RT in exon 7 of the factor VIII gene.";
 RL Am. J. Hum. Genet. 42:867-871(1988).
 RN [16]
 RP VARIANT CYS-1708.
 RX MEDLINE-89274393; PubMed-2499363;
 RA O'Brien D.P., Tuddenham E.G.;
 RT "Purification and characterization of factor VIII 1,689-Cys: a
 RT nonfunctional cofactor occurring in a patient with severe hemophilia
 RL Blood 73:2117-2122(1989).
 RN [17]
 RP VARIANT CYS-391.
 RX MEDLINE-90001543; PubMed-2506948;
 RA Shima M., Ware J., Yoshio A., Fukui H., Fulcher C.A.;
 RT "An arginine to cysteine amino acid substitution at a critical
 RT thrombin cleavage site in a dysfunctional factor VIII molecule.";
 RL Blood 74:1612-1617(1989).
 RN [18]
 RP VARIANT LEU-189.
 RX MEDLINE-90057680; PubMed-2510835;
 RA Chan V., Chan T.K., Tong T.M., Todd D.;
 RT "A novel missense mutation in exon 4 of the factor VIII:C gene
 RT resulting in moderately severe hemophilia A.";
 RL Blood 74:2688-2691(1989).
 RN [19]
 RP VARIANT LEU-2326.
 RX MEDLINE-89197216; PubMed-2495245;
 RA Inaba H., Fujimaki M., Kazazian H.H. Jr., Antonarakis S.E.;
 RT "Wild hemophilia A resulting from Arg-to-Leu substitution in exon 26
 RT of the factor VIII gene.";
 RL Hum. Genet. 81:335-338(1989).
 RN [20]
 RP VARIANT HIS-391.
 RX MEDLINE-89264602; PubMed-2498882;
 RA Arai M., Inaba H., Higuchi M., Antonarakis S.E., Kazazian H.H. Jr.,
 RA Fujimaki M., Hoyer L.W.;
 RT "Direct characterization of factor VIII in plasma: detection of a
 RT mutation altering a thrombin cleavage site
 RT (arginine-372-->histidine).";
 RL Proc. Natl. Acad. Sci. U.S.A. 86:4277-4281(1989).
 RN [21]
 RP VARIANT CYS-1708.
 RX MEDLINE-90105723; PubMed-2104766;
 RA Arai M., Higuchi M., Antonarakis S.E., Kazazian H.H. Jr.,
 RA Phillips J.A. III, Janco R.L., Hoyer L.W.;
 RT "Characterization of a thrombin cleavage site mutation (Arg 1689 to
 RT Cys) in the factor VIII gene of two unrelated patients with
 RT cross-reacting material-positive hemophilia A.";
 RL Blood 75:384-389(1990).
 RN [22]
 RP VARIANTS GLN-2228 AND LEU-2326.
 RX MEDLINE-90123183; PubMed-2105106;
 RA Casula L., Murru S., Pecorara M., Ristaldi M.S., Restagno G.,
 RA Mancuso G., Morfini M., de Biasi R., Baudo F., Carbonara A.;
 RT "Recurrent mutations and three novel rearrangements in the factor
 RT VIII gene of hemophilia A patients of Italian descent.";
 RL Blood 75:662-670(1990).
 RN [23]
 RP VARIANT CYS-391.
 RX MEDLINE-90329422; PubMed-1973901;
 RA Pattinson J.K., Mcvey J.H., Boon M., Ajani A., Tuddenham E.G.;
 RT "CRM+ hemophilia A due to a missense mutation (372-->Cys) at the
 RT internal heavy chain thrombin cleavage site.";
 RL Br. J. Haematol. 75:73-77(1990).
 RN [24]
 RP VARIANTS PHE-1699 AND CYS-1708.
 RX MEDLINE-90152691; PubMed-2105906;
 RA Higuchi M., Wong C., Kochhan L., Olek K., Aronis S., Kasper C.K.,
 RA Kazazian H.H., Antonarakis S.E.;
 RT "Characterization of mutations in the factor VIII gene by direct
 RT sequencing of amplified genomic DNA.";
 RL Genomics 6:65-71(1990).
 RN [25]
 RP VARIANTS CYS-1728 AND ASP-1941.
 RX MEDLINE-90159988; PubMed-2106480;
 RA Traystman M.D., Higuchi M., Kasper C.K., Antonarakis S.E.,
 RA Kazazian H.H.;
 RT "Use of denaturing gradient gel electrophoresis to detect point
 RT mutations in the factor VIII gene."

```
Query Match          93.4%; Score 127; DB 1; Length 2351;
Best Local Similarity 96.0%; Pred. No. 5.8e-11;
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 RPNYSRLPGVKVHLKDFPILPGEI 25
   || ||||| ||||| ||||| |||||
Db 503 RPLYSRLPGVKVHLKDFPILPGEI 527

RESULT 2
FAB_MOUSE
ID FAB_MOUSE STANDARD; PRT: 2319 AA.
AC Q06194;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DE 15-JUN-2002 (Rel. 41, Last annotation update)
DE Coagulation factor VIII precursor (Procoagulant component).
GN F8 OR F8C.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RC STRAIN=C57BL/6 X CBA; TISSUE=Liver;
RX MEDLINE=93300511; PubMed=8314577;
RA Elder B., Lakich D., Gletscher J.;
RL *Sequence of the murine factor VIII cDNA.*;
RL Genomics 16:374-379(1993).
CC -!- FUNCTION: FACTOR VIII, ALONG WITH CALCIUM AND PHOSPHOLIPID, ACTS
CC AS A COFACTOR FOR FACTOR IXA WHEN IT CONVERTS FACTOR X TO THE
CC ACTIVATED FORM, FACTOR XA.
CC -!- SUBCELLULAR LOCATION: Extracellular.
CC -!- TISSUE SPECIFICITY: FOUND IN MOST TISSUES.
CC -!- SIMILARITY: CONTAINS 3 F5/8 TYPE A DOMAINS; EACH IS COMPOSED OF
CC 2 PLASTOCYANIN-LIKE REPEATS.
CC -!- SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS.
CC -!- SIMILARITY: STRONG, TO COAGULATION FACTOR V.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; L05573; AAA37385.1; -.
DR PIR; A47004; A47004.
DR HSSP; P00451; LCEG.
DR MGD; MGI:88383; F8.
DR InterPro; IPR001117; Cu-oxidase.
DR InterPro; IPR000421; FA58C.
DR Pfam; PF00394; Cu-oxidase; 3.
DR Pfam; PF00754; F5_F8_type_C; 2.
DR SMART; SM00231; FA58C; 2.
DR PROSITE; PS00079; MULTICOPPER_OXIDASE1; 3.
DR PROSITE; PS01285; FA58C_1; 2.
DR PROSITE; PS01286; FA58C_2; 2.
DR Blood coagulation; Repeat; Plasma; Acute phase; Calcium;
KW Signal; Glycoprotein; Sulfation.
FT SIGNAL 1 19 POTENTIAL.
FT CHAIN 20 2319 COAGULATION FACTOR VIII.
FT DOMAIN 20 349 F5/8 TYPE A 1.
FT DOMAIN 20 199 PLASTOCYANIN-LIKE 1.
FT DOMAIN 207 349 PLASTOCYANIN-LIKE 2.
FT DOMAIN 399 730 F5/8 TYPE A 2.
FT DOMAIN 399 573 PLASTOCYANIN-LIKE 3.
FT DOMAIN 583 730 PLASTOCYANIN-LIKE 4.
FT DOMAIN 760 1640 B.
FT DOMAIN 1683 2008 F5/8 TYPE A 3.
FT DOMAIN 1683 1845 PLASTOCYANIN-LIKE 5.
```

```
FT DOMAIN 1855 2008 PLASTOCYANIN-LIKE 6.
FT DOMAIN 2009 2156 F5/8 TYPE C 1.
FT DOMAIN 2161 392 F5/8 TYPE C 2.
FT SITE 391 392 CLEAVAGE (BY THROMBIN) (BY SIMILARITY).
FT SITE 759 760 CLEAVAGE (BY THROMBIN) (BY SIMILARITY).
FT SITE 1678 1679 CLEAVAGE (BY THROMBIN) (BY SIMILARITY).
FT SITE 1324 1325 CLEAVAGE (ACTIVATION) (BY SIMILARITY).
FT SITE 1640 1641 CLEAVAGE (ACTIVATION) (BY SIMILARITY).
FT MOD_RES 367 367 SULFATION (BY SIMILARITY).
FT MOD_RES 737 737 SULFATION (BY SIMILARITY).
FT MOD_RES 738 738 SULFATION (BY SIMILARITY).
FT MOD_RES 742 742 SULFATION (BY SIMILARITY).
FT MOD_RES 1669 1669 SULFATION (REQUIRED FOR VWF BINDING)
FT MOD_RES 1669 (BY SIMILARITY).
FT MOD_RES 1687 1687 SULFATION (BY SIMILARITY).
FT DISULFID 173 199 PROBABLE.
FT DISULFID 547 573 PROBABLE.
FT DISULFID 1819 1845 BY SIMILARITY.
FT DISULFID 2008 2156 BY SIMILARITY.
FT DISULFID 2161 2313 BY SIMILARITY.
FT CARBOHYD 61 61 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 233 233 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 259 259 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 423 423 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 601 601 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 880 880 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 958 958 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1015 1015 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1022 1022 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1044 1044 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1076 1076 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1087 1087 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1136 1136 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1161 1161 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1192 1192 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1255 1255 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1268 1268 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1273 1273 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1274 1274 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1302 1302 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1316 1316 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1340 1340 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1378 1378 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1797 1797 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2105 2105 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 2319 AA; 266148 MW; FD054DE051DB2A01 CRC64;
```

```
Query Match          69.1%; Score 94; DB 1; Length 2319;
Best Local Similarity 66.7%; Pred. No. 6.1e-06;
Matches 16; Conservative 5; Mismatches 3; Indels 0; Gaps 0;
```

```
QY 2 PNYSRRLPGVKVHLKDFPILPGEI 25
   |::|||::|||::|||::|||
Db 504 PLHARRLRGKIKVKKDLPIHPGEI 527

RESULT 3
FAB_PIG
ID FAB_PIG STANDARD; PRT: 2133 AA.
AC P12263; Q95243;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Coagulation factor VIII precursor (Procoagulant component).
GN F8 OR CF8.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1]
RA HEALEY J.F., LUBIN I.M., LOLLAR P.;
```

Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.

[2]

SEQUENCE OF 705-1573 FROM N.A.

MEDLINE=86287369; PubMed=3016730;

Toole J.J., Pittman D.D., Orr E.C., Murtha P., Wasley L.C., Kaufman R.J.;

"A large region (approximately equal to 95 kDa) of human factor VIII is dispensable for in vitro procoagulant activity.";

Proc. Natl. Acad. Sci. U.S.A. 83:5939-5942(1986).

[3]

SEQUENCE OF 392-759 FROM N.A.

MEDLINE=94179260; PubMed=7510693;

Lubin I.M., Healey J.F., Scandella D., Runge M.S., Lollar P.;

"Elimination of a major inhibitor epitope in factor VIII.";

J. Biol. Chem. 269:8639-8641(1994).

-1- FUNCTION: FACTOR VIII, ALONG WITH CALCIUM AND PHOSPHOLIPID, ACTS AS A COFACTOR FOR FACTOR IXA WHEN IT CONVERTS FACTOR X TO THE ACTIVATED FORM, FACTOR XA.

-1- SUBCELLULAR LOCATION: Extracellular.

-1- SIMILARITY: CONTAINS 3 F5/8 TYPE A DOMAINS; EACH IS COMPOSED OF 2 PLASTOCYANIN-LIKE REPEATS.

-1- SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS.

-1- SIMILARITY: STRONG, TO COAGULATION FACTOR V.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL: U49517; AAB06705.1; -.

PIR: A25945; A25945.

HSP: P00451; 1CFG.

InterPro: IPR001117; Cu-oxidase.

InterPro: IPR000421; FA58_C.

Pfam: PF00394; Cu-oxidase; 3.

Pfam: PF00754; F5_F8_type_C; 2.

SMART: SM00231; FA58C; 2.

PROSITE: PS00079; MULTICOPPER_OXIDASE1; 3.

PROSITE: PS01285; FA58C_1; 2.

PROSITE: PS01286; FA58C_2; 2.

Blood coagulation; Repeat; Plasma; Acute phase; Calcium; Signal; Glycoprotein; Sulfation.

SIGNAL 1 19 POTENTIAL.

CHAIN 20 2133 COAGULATION FACTOR VIII.

FT F5/8 TYPE A 1.

FT PLASTOCYANIN-LIKE 1.

FT PLASTOCYANIN-LIKE 2.

FT F5/8 TYPE A 2.

FT PLASTOCYANIN-LIKE 3.

FT PLASTOCYANIN-LIKE 4.

FT B.

FT F5/8 TYPE A 3.

FT PLASTOCYANIN-LIKE 5.

FT PLASTOCYANIN-LIKE 6.

FT F5/8 TYPE C 1.

FT F5/8 TYPE C 2.

FT CLEAVAGE (BY THROMBIN) (BY SIMILARITY).

FT CLEAVAGE (BY THROMBIN) (BY SIMILARITY).

FT CLEAVAGE (ACTIVATION) (BY SIMILARITY).

FT CLEAVAGE (BY THROMBIN) (BY SIMILARITY).

FT SULFATION (BY SIMILARITY).

FT SULFATION (BY SIMILARITY).

FT SULFATION (BY SIMILARITY).

FT PROBABLE.

FT PROBABLE.

FT PROBABLE.

FT BY SIMILARITY.

FT BY SIMILARITY.

FT BY SIMILARITY.

FT N-LINKED (GLCNAC. .) (POTENTIAL).

FT N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD

FT CARBOHYD 601 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 929 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 985 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 1025 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 1111 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 1181 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 1208 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 1245 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 1265 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 1335 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 1408 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 1611 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 1919 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CONFLICT 713 713 I -> T (IN REF. 2).

FT CONFLICT 734 734 I -> M (IN REF. 2).

FT CONFLICT 792 792 E -> Q (IN REF. 2).

FT CONFLICT 1133 1133 E -> F (IN REF. 2).

FT CONFLICT 1191 1191 I -> L (IN REF. 2).

FT CONFLICT 1209 1209 R -> F (IN REF. 2).

FT CONFLICT 1437 1437 C -> G (IN REF. 2).

FT CONFLICT 1456 1456 F -> R (IN REF. 2).

FT CONFLICT 1539 1539 F -> R (IN REF. 2).

FT CONFLICT 1546 1546 Q -> N (IN REF. 2).

SQ SEQUENCE 2133 AA; 239304 MW; 152BBA8997F570DA CRC64;

Query Match 55.1%; Score 75; DB 1; Length 2133;

Best Local Similarity 83.3%; Pred. No. 0.0044;

Matches 15; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 7 RLPKGVKHLKDPILPGE 24

DB 509 RLLKGWKLKDPILPGE 526

II II IIIII IIIII

RESULT 4

ID FUT4_HUMAN STANDARD; PRT; 405 AA.

AC P22083;

DT 01-AUG-1991 (Rel. 19, Created)

DT 01-OCT-1996 (Rel. 34, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Alpha-(1,3)-fucosyltransferase (EC 2.4.1.-) (Galactoside 3-L-fucosyltransferase) (Fucosyltransferase 4) (FUCT-IV) (ELAM-1 ligand DE fucosyltransferase).

DE fucosyltransferase).

GN FUT4 OR ELFT.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Peripheral blood leukocytes;

RX MEDLINE=91373370; PubMed=1716630;

RA Lowe J.B., Kukowska-Latallo J.F., Nair R.P., Larsen R.D., Marks R.M., Macher B.A., Kelly R.J., Ernst L.K.;

RT "Molecular cloning of a human fucosyltransferase gene that determines expression of the Lewis x and VIM-2 epitopes but not ELAM-1-dependent cell adhesion.";

RT J. Biol. Chem. 266:17467-17477(1991).

RL [2]

RN SEQUENCE FROM N.A.

RP SEQUENCE FROM N.A.

RX MEDLINE=91084863; PubMed=1702034;

RA Goelz S.E., Hession C., Goff D., Griffiths B., Tizard R., Newman B., Chi-Rosso G., Lobb R.;

RT "ELFT: a gene that directs the expression of an ELAM-1 ligand.";

RL Cell 63:1349-1356(1990).

RN [3]

RP SEQUENCE OF 1-400 FROM N.A.

RX MEDLINE=92042084; PubMed=1718983;

RA Kumar R., Potvin B., Muller W.A., Stanley P.;

RT "Cloning of a human alpha(1,3)-fucosyltransferase gene that encodes ELFT but does not confer ELAM-1 recognition on Chinese hamster ovary cell transfectants.";

```

RL J. Biol. Chem. 266:21777-21783(1991).
CC -|- FUNCTION: MAY CATALYSE ALPHA-1,3 GLYCOSIDIC LINKAGES INVOLVED IN
CC THE EXPRESSION OF LEWIS X/SSEA-1 AND VIM-2 ANTIGENS.
CC -|- PATHWAY: Glycosylation.
CC -|- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. MEMBRANE-BOUND
CC FORM IN TRANS CISTERNAE OF GOLGI.
CC -|- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 10.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M65030; AAA92977.1; -.
DR EMBL; M58596; AAA63172.1; -.
DR EMBL; M58597; AAA63173.1; ALT_INIT.
DR EMBL; S65161; AAB20349.1; -.
DR PIR; A36340; A36340.
DR Genew; HGNC:4015; FUT4.
DR MIM; 104230; -.
DR InterPro; IPR001503; GT_10.
DR Pfam; PF00852; Glyco_transf.10; 1.
DR Transfaser; Glycosyltransferase; Transmembrane; Glycoprotein;
KW Signal-anchor; Golgi stack.
FT DOMAIN 1 22 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 23 47 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT FT (POTENTIAL).
FT DOMAIN 48 405 LUMENAL, CATALYTIC (POTENTIAL).
FT CARBOHYD 91 91 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 190 190 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 87 87 P -> R (IN REF. 2 AND 3).
FT CONFLICT 241 241 E -> D (IN REF. 3).
FT SEQUENCE 405 AA; 45569 MW; DE72EIFDC390268D CRC64;

Query Match 35.0%; Score 48.5; DB 1; Length 405;
Best Local Similarity 50.0%; Pred. No. 7.2;
Matches 10; Conservative 3; Mismatches 5; Indels 1; Gaps 1;

OY 1 RPNYSRRLPKGV-KHLKDPF 19
| | | | | | | | | |
DB 319 RANVERFVPRGAFIHVDFFP 338

RESULT 5
FUT4_MOUSE
ID FUT4_MOUSE STANDARD; PRT; 433 AA.
AC Q11127;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Alpha-(1,3)-fucosyltransferase (EC 2.4.1.-) (Galactoside 3-L-
DE fucosyltransferase) (Fucosyltransferase 4) (FUCT-IV).
GN FUT4 OR ELFT.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=9602767; PubMed=7559635;
RA Gersten K.M., Natsuka S., Trinchera M., Petryniak B., Kelly R.J.,
RA Hiraiwa N., Jenkins N.A., Gilbert D.J., Copeland N.G., Lowe J.B.;
RT "Molecular cloning, expression, chromosomal assignment, and tissue-
RT specific expression of a murine alpha-(1,3)-fucosyltransferase locus
RT corresponding to the human ELAM-1 ligand fucosyl transferase.";
RL J. Biol. Chem. 270:25047-25056(1995).
RL [2]
RP SEQUENCE FROM N.A. (SHORT FORM).
RC STRAIN=129/sv; TISSUE=Liver;
RX MEDLINE=97037075; PubMed=8882722;

Ozawa M., Muramatsu T.;
RT "Molecular cloning and expression of a mouse alpha-1,3
RT fucosyltransferase gene that shows homology with the human alpha-1,3
RT fucosyltransferase IV gene.";
RL J. Biochem. 119:302-308(1996).
CC -|- FUNCTION: MAY CATALYSE ALPHA-1,3 GLYCOSIDIC LINKAGES INVOLVED IN
CC THE EXPRESSION OF LEWIS X/SSEA-1 AND VIM-2 ANTIGENS.
CC -|- PATHWAY: Glycosylation.
CC -|- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. MEMBRANE-BOUND
CC FORM IN TRANS CISTERNAE OF GOLGI.
CC -|- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG FORM (SHOWN HERE) AND A
CC SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -|- TISSUE SPECIFICITY: HIGHEST EXPRESSION IN STOMACH AND COLON.
CC IT ALSO EXPRESSED IN THE LUNG, TESTIS, UTERUS, SMALL INTESTINE
CC AND TO A LESSER EXTENT IN SPLEEN, AND OVARY. PRESENT IN TRACE
CC AMOUNTS IN BRAIN, THYMUS, HEART, SMOOTH MUSCLE, KIDNEY AND BONE
CC NARROW. NOT FOUND IN LIVER, SALIVARY GLAND AND PANCREAS.
CC -|- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 10.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U33457; AAC52269.1; -.
DR EMBL; D63380; BAA09697.1; -.
DR EMBL; D63379; BAA09696.1; -.
DR MGD; MGI:95594; Fut4.
DR InterPro; IPR001503; GT_10.
DR Pfam; PF00852; Glyco_transf.10; 1.
DR Transfaser; Glycosyltransferase; Transmembrane; Glycoprotein;
KW Signal-anchor; Golgi stack; Alternative splicing.
FT DOMAIN 1 52 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 53 74 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT FT (POTENTIAL).
FT DOMAIN 75 433 LUMENAL, CATALYTIC (POTENTIAL).
FT CARBOHYD 117 117 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 218 218 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARSPLIC 1 33 MISSING (IN SHORT ISOFORM).
FT CONFLICT 252 252 Q -> P (IN REF. 2).
FT CONFLICT 257 257 R -> Q (IN REF. 2).
FT CONFLICT 260 260 V -> E (IN REF. 2).
FT CONFLICT 273 273 R -> Q (IN REF. 2).
FT SEQUENCE 433 AA; 49481 MW; 2401822F02B5D021 CRC64;

Query Match 35.7%; Score 48.5; DB 1; Length 433;
Best Local Similarity 50.0%; Pred. No. 7.8;
Matches 10; Conservative 3; Mismatches 6; Indels 1; Gaps 1;

OY 1 RPNYSRRLPKGV-KHLKDPF 19
| | | | | | | | | |
DB 347 RANVERFVPRGAFIHVDFFP 366

RESULT 6
DPO2_RAT
ID DPO2_RAT STANDARD; PRT; 495 AA.
AC O89043;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE DNA polymerase alpha 70 kDa subunit (DNA polymerase subunit B)
DE (DNA polymerase subunit II) (fragment).
GN POLA2.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.

```

```

RC STRAIN-Sprague-Dawley; TISSUE=Liver;
RA Popanda O., Flohr C., Thielmann H.W.;
RT "A mutation in the gene of subunit II of DNA polymerase alpha from
RT Novikoff cells is concomitant with altered physico-chemical
RT properties of the enzyme.";
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: MAY PLAY AN ESSENTIAL ROLE AT THE EARLY STAGE OF
CC CHROMOSOMAL DNA REPLICATION BY COUPLING THE POLYMERASE
CC ALPHA/PRIMASE COMPLEX TO THE CELLULAR REPLICATION MACHINERY (BY
CC SIMILARITY).
CC -1- SUBUNIT: DNA POLYMERASE ALPHA-PRIMASE IS A FOUR SUBUNIT ENZYME
CC (SUBUNITS A, B, C AND D), WHICH IS ASSEMBLED THROUGHOUT THE CELL
CC CYCLE. THE LARGEST SUBUNIT (SUBUNIT A) HAS DNA POLYMERASE
CC ACTIVITY, THE TWO SMALLEST SUBUNITS (SUBUNITS C AND D) HAVE DNA
CC PRIMASE ACTIVITY. SUBUNIT B BINDS TO SUBUNIT A.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- PTM: PHOSPHORYLATED IN A CELL CYCLE-DEPENDENT MANNER, IN G2/M
CC PHASE (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE DNA POLYMERASE ALPHA SUBUNIT B FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AJ011606; CAA09721.1; -.
CC DNA replication; Nuclear protein; Phosphorylation.
CC KW D_NTER 1 1
CC FT DOMAIN 10 52 PRO/SER/THR-RICH (HYDROPHILIC).
CC SEQUENCE 495 AA; 54952 MW; 1816F035737C4A0B CRC64;
CC -----
Query Match 35.3%; Score 48; DB 1; Length 495;
Best Local Similarity 36.8%; Pred. No. 11;
Matches 7; Conservative 6; Mismatches 6; Indels 0; Gaps 0;
QY 7 RLPGVKHLKDFPILPGEI 25
:| | | | | | | | |
DB 178 QIPVDVSELKDYSLFPGQV 196
-----
RESULT 7
DPO2_MOUSE STANDARD; PRT; 600 AA.
AC P33611;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE DNA polymerase alpha 70 kDa subunit (DNA polymerase subunit B).
DE POLA2.
GN Mus musculus (Mouse).
OS Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 84-102; 269-285 AND 394-403.
RX MEDLINE-93216788; PubMed=8463324;
RA Miyazawa H., Izumi M., Tada S., Takada R., Masutani M., Uli M.,
RA Hanaoka F.;
RT "Molecular cloning of the cDNAs for the four subunits of mouse DNA
RT polymerase alpha-prime complex and their gene expression during
RT cell proliferation and the cell cycle.";
RL J. Biol. Chem. 268:8111-8123(1993).
CC -1- FUNCTION: MAY PLAY AN ESSENTIAL ROLE AT THE EARLY STAGE OF
CC CHROMOSOMAL DNA REPLICATION BY COUPLING THE POLYMERASE
CC ALPHA/PRIMASE COMPLEX TO THE CELLULAR REPLICATION MACHINERY (BY
CC SIMILARITY).
CC -1- SUBUNIT: DNA POLYMERASE ALPHA-PRIMASE IS A FOUR SUBUNIT ENZYME
CC (SUBUNITS A, B, C AND D), WHICH IS ASSEMBLED THROUGHOUT THE CELL
CC CYCLE. THE LARGEST SUBUNIT (SUBUNIT A) HAS DNA POLYMERASE
CC ACTIVITY, THE TWO SMALLEST SUBUNITS (SUBUNITS C AND D) HAVE DNA

```

```

CC PRIMASE ACTIVITY. SUBUNIT B BINDS TO SUBUNIT A.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- PTM: PHOSPHORYLATED IN A CELL CYCLE-DEPENDENT MANNER, IN G2/M
CC PHASE (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE DNA POLYMERASE ALPHA SUBUNIT B FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: D13546; BAA02746.1; -.
CC PIR: B46642; B46642.
CC MGD; MGI:99690; Pola2.
CC DNA replication; Nuclear protein; Phosphorylation.
CC KW DOMAIN 101 107 POLY-GLU.
CC FT DOMAIN 115 157 PRO/SER/THR-RICH (HYDROPHILIC).
CC SEQUENCE 600 AA; 66267 MW; 79F94BE6EF33FEB3 CRC64;
CC -----
Query Match 35.3%; Score 48; DB 1; Length 600;
Best Local Similarity 26.1%; Pred. No. 13;
Matches 6; Conservative 9; Mismatches 8; Indels 0; Gaps 0;
QY 3 NYSRRLPKGVKHLKDFPILPGEI 25
:| | | | | | | | |
DB 279 SYGAQIPVDLSSELKDYSLFPGQV 301
-----
RESULT 8
GLVD_SCHPO STANDARD; PRT; 467 AA.
ID GLVD_SCHPO
AC O13972;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Probable serine hydroxymethyltransferase, cytosolic (EC 2.1.2.1)
DE (Serine methylase) (Glycine hydroxymethyltransferase) (SHMT).
GN SPAC24C9.12C.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomyces.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=972;
RX MEDLINE-21848401; PubMed=11859360;
RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA James K., Jones M., Jones M., Leather S., McDonald S., McLean J.,
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odeh C.,
RA Oliver K., O'Neil S., Pearson D., Quail M.A., Rabinowitsch E.,
RA Rutherford K., Rutter S., Saunders R., Seeger K., Sharp S.,
RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,
RA Weltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,
RA Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,
RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA Cerrutti L., Lowe T., McComble W.R., Paulsen I., Potashkin J.,

```



```

RA Shpakovskii G.V., Ussery D., Barrell B.G., Nurse P.;
RT "The genome sequence of Schizosaccharomyces pombe.";
RL Nature 415:871-880(2002).
CC -1- FUNCTION: INTERCONVERSION OF SERINE AND GLYCINE.
CC -1- CATALYTIC ACTIVITY: 5,10-methylenetetrahydrofolate + glycine +
CC H(2)O = tetrahydrofolate + L-serine.
CC -1- COFACTOR: PYRIDOXAL PHOSPHATE (BY SIMILARITY).
CC -1- PATHWAY: KEY ENZYME IN THE BIOSYNTHESIS OF PURINES, LIPIDS,
CC HORMONES AND OTHER COMPONENTS.
CC -1- SUBUNIT: HOMOTETRAMER (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Potential).
CC -1- SIMILARITY: BELONGS TO THE SHMT FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: Z98601; CAB11269.1; -.
CC HSPSP; P07511; ICJ0.
CC InterPro: IPR001085; Gly_HyMetTransf.
CC Pfam: PF00464; SHMT; 1.
CC PROSITE: PS00096; SHMT; 1.
CC Transferrase: Pyridoxal phosphate: One-carbon metabolism.
KW BINDING 243 243 PYRIDOXAL PHOSPHATE (BY SIMILARITY).
FT DOMAIN 238 241 POLY-THR.
FT SEQUENCE 467 AA; 51861 MW; 0C21D7EF010C3725 CRC64;
SQ
Query Match 33.8%; Score 46; DB 1; Length 467;
Best Local Similarity 45.5%; Pred. No. 20;
Matches 10; Conservative 1; Mismatches 11; Indels 0; Gaps 0;

QY 3 NYSRLPKGVKHLKDFPILPGE 24
I : III IIII I
Db 419 NLQELPKDANKLKDFRAKLGE 440

RESULT 9
UBIG_XYLFA
ID UBIG_XYLFA STANDARD; PRT; 246 AA.
AC Q9PAM5;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE 3-demethylubiquinone-9 3-methyltransferase (EC 2.1.1.64) (3,4-
DE dihydroxy-5-hexaprenylbenzoate methyltransferase) (DHMB
DE methyltransferase).
GN UBIG OR XP2471.
OS Xylella fastidiosa.
OC Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;
OC Xylella.
OX NCBI_TaxID=2371;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=9a5c;
RX MEDLINE=20365717; PubMed=10910347;
RA Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,
RA Alvaranga R., Alves J.E., Araya J.E., Baia G.S., Baptista C.S.,
RA Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,
RA Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carrer H.,
RA Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,
RA Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorry H.,
RA Pacinanci A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,
RA Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furian L.R.,
RA Garnier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,
RA Ho P.L., Hohelsel J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,
RA Krieger J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite L.C.C.,
RA Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,
RA Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Marino C.L.,
RA Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,

```

```

RA Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,
RA Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,
RA Nhani A. Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,
RA de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,
RA Peixoto B.R., Pereira G.A.G., Pereira H.A. Jr., Pesquero J.B.,
RA Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,
RA de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,
RA da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,
RA da Silva J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,
RA de Souza A.P., Terenzi M.F., Truffi D., Tsal S.M., Tshako M.H.,
RA Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.B.,
RA Zago M.A., Zatz M., Meidanis J., Setubal J.C.;
RT "The genome sequence of the plant pathogen Xylella fastidiosa.";
RL Nature 406:151-159(2000).
CC -1- CATALYTIC ACTIVITY: S-adenosyl-L-methionine + 3-
CC demethylubiquinone-9 - S-adenosyl-L-homocysteine + ubiquinone-9.
CC -1- PATHWAY: Ubiquinone biosynthesis.
CC -1- SIMILARITY: BELONGS TO THE UBIG/COQ3 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AF004055; AAF85269.1; -.
CC InterPro: IPR001601; Methyltransf.
CC InterPro: IPR000051; SAM_bind.
CC Ubiquinone biosynthesis; Transferase; Methyltransferase;
KW Complete proteome.
SQ SEQUENCE 246 AA; 27121 MW; 5212107D6363D3DF CRC64;

Query Match 33.5%; Score 45.5; DB 1; Length 246;
Best Local Similarity 55.0%; Pred. No. 12;
Matches 11; Conservative 1; Mismatches 7; Indels 1; Gaps 1;

QY 6 RRLPKGVKHLKDFPILPGEI 25
I : IIII IIII I
Db 178 RLPLKGTHTQYKDF-IRPAEL 196

RESULT 10
MHPA_ECOLI
ID MHPA_ECOLI STANDARD; PRT; 554 AA.
AC P77337; P71203; P77047;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE 3-(3-hydroxy-phenyl)propionate hydroxylase (EC 1.14.13.-).
GN MHPA OR B0347.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / W3110;
RA Kawamukai M.;
RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.
[2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / CS520;
RA Ferrandez A., Garcia J.L., Diaz E.;
RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
[3]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,

```

```

RA  Mau B., Shao Y.;
RT  "The complete genome sequence of Escherichia coli K-12.";
RL  Science 277:1453-1474(1997).
RN  [4]
RP  SEQUENCE FROM N.A.
RC  STRAIN-K12 / MGI1655;
RA  Duncan M., Allen E., Araujo R., Aparicio A.M., Chung E., Davis K.,
RA  Federspiel N., Hyman R., Kalman S., Komp C., Kurdi O., Lew H., Lin D.,
RA  Namath A., Oefner P., Roberts D., Schramm S., Davis R.W.;
RL  Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
CC  -|- COFACTOR: FAD (BY SIMILARITY).
CC  -|- PATHWAY: 3-hydroxyphenylpropanate degradation.
CC  -|- SIMILARITY: BELONGS TO THE PHEA/TFDB FAMILY OF FAD MONOOXYGENASES.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
CC  EMBL: D86239; BAA13052.1; -
CC  EMBL: Y09555; CAA70747.1; -
CC  EMBL: AE000142; AAC73450.1; -
CC  EMBL: U73857; AAB18071.1; -
CC  EcoGene; EG20273; mhpa.
CC  InterPro: IPR000733; Flav_monooxygenase.
CC  InterPro: IPR002938; Moxv_FAD_binding.
CC  InterPro: IPR000205; NAD_binding.
CC  InterPro: IPR003042; Rng_monooxygenase.
CC  Pfam: PF01360; Monooxygenase; 1.
CC  Pfam: PF01494; FAD_binding_3; 1.
CC  PRINTS; PR00420; RINGMONOXNASE.
CC  Aromatic hydrocarbons catabolism; Oxidoreductase; Flavoprotein; FAD;
KW  Complete proteome.
FT  NP_BIND 17 46 FAD (POTENTIAL).
FT  NP_BIND 285 295 FAD (POTENTIAL).
FT  CONFLICT 272 272 Q -> H (IN REF. 1).
FT  CONFLICT 360 360 L -> P (IN REF. 1).
FT  CONFLICT 502 502 W -> G (IN REF. 1).
SQ  SEQUENCE 554 AA; 62185 MW; 1D56CB799E9F8A8E CRC64;

Query Match 33.5%; Score 45.5; DB 1; Length 554;
Best Local Similarity 41.7%; Pred. No. 30;
Matches 10; Conservative 4; Mismatches 9; Indels 1; Gaps 1;

OY 1 RPNVSRRLPKGVKHLKDFEPLTIGE 24
DB 220 RPYVSAALPHAVRRF-EFWMVPE 242

RESULT 11
MURA_SYNY3
ID MURA_SYNY3 STANDARD; PRT; 438 AA.
AC Q55673;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE UDP-N-acetylglucosamine 1-carboxyvinyltransferase (EC 2.5.1.7)
DE (Enolpyruvate transferase) (UDP-N-acetylglucosamine enolpyruvyl
DE transferase) (EPT).
GN MURA OR MURZ OR SLR0017.
OS Synecocystis sp. (strain PCC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; Synecocystis.
OX NCBI_TaxID=1148;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96127529; PubMed=8590279;
RA Kaneko T., Tanaka A., Sato S., Kotani H., Sazuka T., Miyajima N.,
RA Sugita M., Tabata S.;
RT "Sequence analysis of the genome of the unicellular cyanobacterium
RT Synecocystis sp. strain PCC6803. I. Sequence features in the 1 Mb

```

```

RT region from map positions 64% to 92% of the genome.";
RL DNA Res 2:153-156(1995).
CC -|- FUNCTION: CELL WALL FORMATION. ADDS ENOLPYRUVYL TO UDP-N-
CC ACETYLGLUCOSAMINE (BY SIMILARITY).
CC -|- CATALYTIC ACTIVITY: Phosphoenolpyruvate + UDP-N-acetyl-D-
CC glucosamine = phosphate + UDP-N-acetyl-3-O-(1-carboxyvinyl)-D-
CC glucosamine.
CC -|- PATHWAY: Peptidoglycan biosynthesis; first step.
CC -|- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC -|- SIMILARITY: BELONGS TO THE EPSP SYNTHASE FAMILY. MURA SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: D64000; BAA10199.1; -
CC HSSP: P33038; IDLG.
CC InterPro: IPR001986; EPSP_synthase.
CC Pfam: PF00275; EPSP_synthase; 1.
CC ProDom: PD001867; EPSP_synthase; 1.
CC TIGRFAMs; TIGR01072; murA; 1.
KW Peptidoglycan synthesis; Cell wall; Cell division; Transferase;
FT ACT_SITE 129 129 BINDS PEP (BY SIMILARITY).
SQ SEQUENCE 438 AA; 46604 MW; C310AB287F2F0F55 CRC64;

Query Match 33.1%; Score 45; DB 1; Length 438;
Best Local Similarity 47.8%; Pred. No. 27;
Matches 11; Conservative 3; Mismatches 7; Indels 2; Gaps 1;

OY 5 SRRLPKGVK--HLKDFPILPGEI 25
DB 228 SRIINGVEKLHSTDFPIPDRI 250

RESULT 12
IT44_XENLA
ID IT44_XENLA STANDARD; PRT; 1032 AA.
AC Q91687; Q06273;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Integrin alpha-4 precursor (Integrin alpha-IV) (VIA-4).
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96379747; PubMed=8787760;
RA Ramos J.W., Whittaker C.A., Desimone D.W.;
RT "Integrin-dependent adhesive activity is spatially controlled by
RT inductive signals at gastrulation.";
RL Development 122:2873-2883(1996).
RN [2]
RP SEQUENCE OF 308-379 FROM N.A.
RX MEDLINE=94008528; PubMed=8404528;
RA Whittaker C.A., Desimone D.W.;
RT "Integrin alpha subunit mRNAs are differentially expressed in early
RT Xenopus embryos.";
RL Development 117:1239-1249(1993).
CC -|- FUNCTION: FIBRONECTIN AND V-CAM ADHESION RECEPTOR.
CC -|- SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA SUBUNIT.
CC -|- SUBCELLULAR LOCATION: TYPE I membrane protein.
CC -|- SIMILARITY: BELONGS TO THE INTEGRIN ALPHA CHAIN FAMILY.
CC -|- SIMILARITY: CONTAINS 7 FG-GAP REPEATS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

```

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMBL; U54497; AAA98673.1; -.
 DR EMBL; L10188; AAA16248.1; -.
 DR HSP; P11215; IABX
 DR InterPro: IPR000413; Integrin_alpha.
 DR Pfam; PF00357; Integrin_A; 1.
 DR Pfam; PF01839; FG-GAP; 5.
 DR PRINTS; PR01185; INTEGRINA.
 DR SMART; SM00191; Int.alpha; 5.
 DR PROSITE; PS00242; INTEGRIN_ALPHA; 1.
 DR Integrin; Cell adhesion; Receptor; Glycoprotein; Transmembrane;
 KW Signal; Repeat.
 FT SIGNAL 1 34 POTENTIAL.
 FT CHAIN . 35 1032 INTEGRIN ALPHA-4.
 FT DOMAIN 35 974 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 975 998 POTENTIAL.
 FT DOMAIN 999 1032 CYTOPLASMIC (POTENTIAL).
 FT REPEAT 50 112 FG-GAP 1.
 FT REPEAT ? ? FG-GAP 2.
 FT REPEAT 197 247 FG-GAP 3.
 FT REPEAT 248 301 FG-GAP 4.
 FT REPEAT 303 361 FG-GAP 5.
 FT REPEAT 364 423 FG-GAP 6.
 FT REPEAT 426 474 FG-GAP 7.
 FT CA_BIND 314 322 POTENTIAL.
 FT CA_BIND 376 384 POTENTIAL.
 FT CA_BIND 438 446 POTENTIAL.
 FT SITE 590 591 CLEAVAGE.
 FT SITE 1001 1005 GPPKR MOTIF.
 FT DISULFID 91 101 BY SIMILARITY.
 FT DISULFID 144 165 BY SIMILARITY.
 FT DISULFID 183 198 BY SIMILARITY.
 FT DISULFID 485 494 BY SIMILARITY.
 FT DISULFID 500 556 BY SIMILARITY.
 FT DISULFID 622 627 BY SIMILARITY.
 FT DISULFID 698 712 BY SIMILARITY.
 FT DISULFID 853 889 BY SIMILARITY.
 FT DISULFID 896 901 BY SIMILARITY.
 FT CARBOHYD 81 81 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 98 98 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 229 229 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 479 479 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 496 496 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 517 517 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 537 537 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 626 626 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 660 660 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 746 746 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 857 857 N-LINKED (GLCNAC. .) (POTENTIAL).
 SQ SEQUENCE 1032 AA; 115215 MW; 6486797D83AAE69E CRC64;

Query Match 33.1%; Score 45; DB 1; Length 1032;
 Best Local Similarity 52.6%; Pred. No. 71;
 Matches 10; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

Qy 7 RLPGVKRHLKDFPILPGEI 25

Db 678 RLPGKLYFKVVDLLEKEI 696

RESULT 13

CA64_HUMAN

ID Q14031;

AC Q14031; STANDARD; PRT; 1678 AA.

DT 01-NOV-1997 (Rel. 35, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Collagen alpha 6(IV) chain precursor.
 GN COL4A6.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_Taxid=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Eye, and kidney;
 RX MEDLINE=94171779; PubMed=8125972;
 RA Ohashi T., Sugimoto M., Mattei M.-G., Ninomiya Y.:
 RT "Identification of a new collagen IV chain, alpha 6(IV), by cDNA
 RT isolation and assignment of the gene to chromosome Xq22, which is the
 RT same locus for COL4A5.";
 RL J. Biol. Chem. 269:7520-7526(1994).
 CC -!- FUNCTION: TYPE IV COLLAGEN IS THE MAJOR STRUCTURAL COMPONENT OF
 CC GLOMERULAR BASEMENT MEMBRANES (GBM), FORMING A 'CHICKEN-WIRE'
 CC MESHWORK TOGETHER WITH LAMININS, PROTEOGLYCANS AND ENACTIN/
 CC NIDOGEN.
 CC -!- SUBUNIT: THERE ARE SIX TYPE IV COLLAGEN ISOFORMS, ALPHA 1(IV)-
 CC ALPHA 6(IV), EACH OF WHICH CAN FORM A TRIPLE HELIX STRUCTURE
 CC WITH 2 OTHER CHAINS TO GENERATE TYPE IV COLLAGEN NETWORK.
 CC -!- SUBCELLULAR LOCATION: CELL SURFACE (POTENTIAL).
 CC -!- DOMAIN: NCP1 CHAINS OF TYPE IV COLLAGEN HAVE A NONCOLLAGENOUS
 CC G-X-Y REPEATS IN THE LONG CENTRAL TRIPLE-HELICAL DOMAIN (WHICH MAY
 CC CAUSE FLEXIBILITY IN THE TRIPLE HELIX), AND A SHORT N-TERMINAL
 CC TRIPLE-HELICAL 7S DOMAIN.
 CC -!- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
 CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
 CC -!- PTM: TYPE IV COLLAGENS CONTAIN NUMEROUS CYSTEINE RESIDUES WHICH
 CC ARE INVOLVED IN INTER- AND INTRAMOLECULAR DISULFIDE BONDING. 12 OF
 CC THESE, LOCATED IN THE NCP1 DOMAIN, ARE CONSERVED IN ALL KNOWN TYPE
 CC IV COLLAGENS.
 CC -!- SIMILARITY: TO OTHER TYPE IV COLLAGENS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

EMBL; D21337; BAA04809.1; -.
 Genew; HGNC:2208; COL4A6.
 MIM; 303631; -.
 InterPro: IPR000087; Collagen.
 InterPro: IPR001442; ProcollagenC4.
 Pfam; PF01391; Collagen; 23.
 Pfam; PF01413; C4; 2.
 ProDom; PD000007; Collagen; 1.
 ProDom; PD003923; ProcollagenC4; 2.
 SMART; SM00111; C4; 2.
 KW Extracellular matrix; Connective tissue; Basement membrane;
 KW Repeat; Hydroxylation; Glycoprotein; Cell adhesion; Collagen; Signal.
 FT SIGNAL 1 15 POTENTIAL.
 FT CHAIN 16 1678 COLLAGEN ALPHA 6(IV) CHAIN.
 FT DOMAIN 16 45 7S DOMAIN.
 FT DOMAIN 46 1450 TRIPLE-HELICAL REGION.
 FT DOMAIN 1451 1678 NONHELICAL REGION (NCP1).
 FT SITE 514 516 CELL ATTACHMENT SITE (POTENTIAL).
 FT SITE 559 561 CELL ATTACHMENT SITE (POTENTIAL).
 FT SITE 985 987 CELL ATTACHMENT SITE (POTENTIAL).
 FT DISULFID 1469 1558 OR 1555 (BY SIMILARITY).
 FT DISULFID 1502 1555 OR 1556 (BY SIMILARITY).
 FT DISULFID 1514 1520 BY SIMILARITY.
 FT DISULFID 1577 1674 OR 1671 (BY SIMILARITY).
 FT DISULFID 1611 1671 OR 1674 (BY SIMILARITY).
 FT DISULFID 1623 1630 BY SIMILARITY.
 FT CARBOHYD 126 126 N-LINKED (GLCNAC. .) (POTENTIAL).
 SQ SEQUENCE 1678 AA; 162461 MW; 65E4EA28D3D37BD CRC64;

```

Query Match      33.1%; Score 45; DB 1; Length 1678;
Best Local Similarity 56.2%; Pred. No. 1.2e+02;
Matches 9; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 9 PKGVKHLKDFPILPGE 24
DB 715 PRGEKGLPGFGLPGK 730

RESULT 14
CA44_HUMAN
ID CA44_HUMAN STANDARD; PRT; 1690 AA.
AC P53420;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Collagen alpha 4(IV) chain precursor.
GN COL4A4.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Kidney;
RX MEDLINE=95014445; PubMed=7523402;
RA Leinonen A., Mariyama M., Mochizuki T., Tryggvason K., Reiders S.T.;
RT "Complete primary structure of the human type IV collagen alpha 4(IV)
RT chain. Comparison with structure and expression of the other alpha
RT (IV) chains.";
RL J. Biol. Chem. 269:26172-26177(1994).
RN [2]
RN SEQUENCE OF 1-23 FROM N.A.
RX MEDLINE=98196854; PubMed=9537505;
RA Momota R., Sugimoto M., Ohashi T., Kigasawa K., Yoshioka H.,
RA Ninomiya Y.;
RT "Two genes, COL4A3 and COL4A4 coding for the human alpha3(IV) and
RT alpha4(IV) collagen chains are arranged head-to-head on chromosome
RT 2q36.";
RL FEBS Lett. 424:11-16(1998).
RN [3]
RP SEQUENCE OF 1219-1690 FROM N.A.
RC TISSUE=Eye;
RX MEDLINE=93374047; PubMed=8365481;
RA Sugimoto M., Ohashi T., Yoshioka H., Matsuo N., Ninomiya Y.;
RT "cDNA isolation and partial gene structure of the human alpha 4(IV)
RT collagen chain.";
RL FEBS Lett. 330:122-128(1993).
RN [4]
RP SEQUENCE OF 1407-1507 FROM N.A.
RX MEDLINE=93054733; PubMed=1429714;
RA Kamagata Y., Mattel M.-G., Ninomiya Y.;
RT "Isolation and sequencing of cDNAs and genomic DNAs encoding the
RT alpha 4 chain of basement membrane collagen type IV and assignment of
RT the gene to the distal long arm of human chromosome 2.";
RL J. Biol. Chem. 267:23753-23758(1992).
RN [5]
RP REVIEW ON VARIANTS.
RX MEDLINE=97338662; PubMed=9195222;
RA Lemmink H.H., Schroeder C.H., Monnens L.A.H., Smeets H.J.M.;
RT "The clinical spectrum of type IV collagen mutations.";
RL Hum. Mutat. 9:477-499(1997).
RN [6]
RP VARIANT AS SER=1201.
RX MEDLINE=95078927; PubMed=7987396;
RA Mochizuki T., Lemmink H.H., Mariyama M., Antignac C., Gubler M.-C.,
RA Pirsion Y., Verellen-Dumoulin C., Chan B., Schroeder C.H.,
RA Smeets H.J.M., Reiders S.T.;
RT "Identification of mutations in the alpha 3(IV) and alpha 4(IV)
RT collagen genes in autosomal recessive Alport syndrome.";
RL Nat. Genet. 8:77-82(1994).
RN [7]
RP VARIANT BFH GUJ-897.

```

```

RX MEDLINE=96379660; PubMed=8787673;
RA Lemmink H.H., Nillesen W.N., Mochizuki T., Schroeder C.H.,
RA Brunner H.G., van Oost B.A., Monnens L.A.H., Smeets H.J.M.;
RT "Benign familial hematuria due to mutation of the type IV collagen
RT alpha4 gene.";
RL J. Clin. Invest. 98:1114-1118(1996).
RN [8]
RP VARIANTS AS, AND VARIANTS.
RX MEDLINE=99011253; PubMed=9792860;
RA Boye E., Mollet G., Forestier L., Cohen-Solal L., Heidet L.,
RA Cochat P., Gruenfeld J.-P., Palcoux J.-B., Gubler M.-C., Antignac C.;
RT "Determination of the genomic structure of the COL4A4 gene and of
RT novel mutations causing autosomal recessive Alport syndrome.";
RL Am. J. Hum. Genet. 63:1329-1340(1998).
CC -!- FUNCTION: TYPE IV COLLAGEN IS THE MAJOR STRUCTURAL COMPONENT OF
CC GLOMERULAR BASEMENT MEMBRANES (GBM), FORMING A 'CHICKEN-WIRE'
CC MESHWORK TOGETHER WITH LAMININS, PROTEOGLYCANS AND ENTACTIN/
CC NIDOGEN.
CC -!- SUBUNIT: THERE ARE SIX TYPE IV COLLAGEN ISOFORMS, ALPHA 1(IV)-
CC ALPHA 6(IV), EACH OF WHICH CAN FORM A TRIPLE HELIX STRUCTURE
CC WITH 2 OTHER CHAINS TO GENERATE TYPE IV COLLAGEN NETWORK.
CC -!- SUBCELLULAR LOCATION: CELL SURFACE (POTENTIAL).
CC -!- TISSUE SPECIFICITY: ALPHA 3 AND ALPHA 4 TYPE IV COLLAGENS ARE
CC COLOCALIZED AND PRESENT ONLY IN BASEMENT MEMBRANES OF KIDNEY, EYE,
CC COCHLEA, LUNG AND BRAIN.
CC -!- DOMAIN: ALPHA CHAINS OF TYPE IV COLLAGEN HAVE A NONCOLLAGENOUS
CC DOMAIN (NC1) AT THEIR C-TERMINUS, FREQUENT INTERRUPTIONS OF THE
CC G-X-Y REPEATS IN THE LONG CENTRAL TRIPLE-HELICAL DOMAIN (WHICH MAY
CC CAUSE FLEXIBILITY IN THE TRIPLE HELIX), AND A SHORT N-TERMINAL
CC TRIPLE-HELICAL 7S DOMAIN.
CC -!- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
CC -!- PTM: TYPE IV COLLAGENS CONTAIN NUMEROUS CYSTEINE RESIDUES WHICH
CC ARE INVOLVED IN INTER- AND INTRAMOLECULAR DISULFIDE BONDING. 12 OF
CC THESE, LOCATED IN THE NC1 DOMAIN, ARE CONSERVED IN ALL KNOWN TYPE
CC IV COLLAGENS.
CC -!- DISEASE: DEFECTS IN COL4A4 ARE A CAUSE OF TYPE II AUTOSOMAL
CC RECESSIVE FORM OF ALPORT SYNDROME (AS), AN HEREDITARY
CC GLOMERULONEPHROPATHY CHARACTERIZED BY PROGRESSIVE RENAL FAILURE,
CC HEMATURIA AND DEAFNESS. THE RECESSIVE FORM OCCURS EQUALLY BETWEEN
CC MALES AND FEMALES.
CC -!- DISEASE: DEFECTS IN COL4A4 ARE A CAUSE OF FAMILIAL BENIGN
CC HEMATURIA (FBH) OR THIN BASEMENT MEMBRANE DISEASE. FBH IS
CC CHARACTERIZED BY PERSISTENT HEMATURIA, AN ELECTRON MICROSCOPICALLY
CC DETECTABLE THIN GLOMERULAR BASEMENT MEMBRANE (GBM) AND AN
CC AUTOSOMAL DOMINANT MODE OF INHERITANCE. RENAL FUNCTION REMAINS
CC NORMAL. IN CHILDREN DIFFERENTIATION BETWEEN FBH AND AS CAN BE
CC DIFFICULT, BECAUSE BOTH DISORDERS ARE MANIFESTED BY PERSISTENT
CC HEMATURIA AND THIN GBM AT THAT AGE.
CC -!- SIMILARITY: TO OTHER TYPE IV COLLAGENS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X81053; CAA56943.1; -.
CC EMBL; AB008496; BAA25065.1; -.
CC EMBL; D17391; BAA04214.1; -.
CC Genew; HGNC:2206; COL4A4.
CC MIM; 120131; -.
CC MIM; 141200; -.
CC MIM; 203780; -.
CC InterPro; IPR000087; Collagen.
CC InterPro; IPR001442; ProcollagenC4.
CC Pfam; PF01391; Collagen; 21.
CC Pfam; PF01413; C4; 2.
CC ProDom; PD000007; Collagen; 1.
CC ProDom; PD003923; ProcollagenC4; 2.
CC SMART; SM00111; C4; 2.

```

Extracellular matrix; Connective tissue; Basement membrane; Repeat;
Hydroxylation; Collagen; Glycoprotein; Signal; Disease mutation;
Polymorphism; Alport syndrome.

FT SIGNAL 1 38 POTENTIAL.
FT CHAIN 39 1690 COLLAGEN ALPHA 4(IV) CHAIN.
FT DOMAIN 39 64 7S DOMAIN.
FT DOMAIN 65 1459 TRIPLE-HELICAL REGION.
FT DOMAIN 1460 1690 NONHELICAL REGION (NCI).
FT SITE 94 96 CELL ATTACHMENT SITE (POTENTIAL).
FT SITE 145 147 CELL ATTACHMENT SITE (POTENTIAL).
FT SITE 189 191 CELL ATTACHMENT SITE (POTENTIAL).
FT SITE 310 312 CELL ATTACHMENT SITE (POTENTIAL).
FT SITE 724 726 CELL ATTACHMENT SITE (POTENTIAL).
FT SITE 785 787 CELL ATTACHMENT SITE (POTENTIAL).
FT SITE 989 991 CELL ATTACHMENT SITE (POTENTIAL).
FT SITE 1206 1206 CLEAVAGE (BY COLLAGENASE)
(BY SIMILARITY).
FT SITE 1212 1214 CELL ATTACHMENT SITE (POTENTIAL).
FT DISULFID 1480 1569 OR 1566 (BY SIMILARITY).
FT DISULFID 1513 1566 OR 1569 (BY SIMILARITY).
FT DISULFID 1525 1531 BY SIMILARITY.
FT DISULFID 1588 1686 OR 1683 (BY SIMILARITY).
FT DISULFID 1622 1683 OR 1686 (BY SIMILARITY).
FT DISULFID 1634 1641 BY SIMILARITY.
FT CARBOHYD 142 142 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 669 669 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARIANT 441 446 MISSING (IN AS).
/FTID=VAR_008148.
FT VARIANT 545 545 G -> A.
/FTID=VAR_008149.
FT VARIANT 570 570 E -> Q.
/FTID=VAR_008150.
FT VARIANT 897 897 G -> E (IN FBH).
/FTID=VAR_001912.
FT VARIANT 931 931 A -> T.
/FTID=VAR_008151.
FT VARIANT 1004 1004 L -> P.
/FTID=VAR_008152.
FT VARIANT 1030 1030 G -> V (IN AS).
/FTID=VAR_008153.
FT VARIANT 1201 1201 G -> S (IN AS).
/FTID=VAR_001913.
FT VARIANT 1402 1402 P -> S.
/FTID=VAR_008154.
FT VARIANT 1572 1572 P -> L (IN AS).
/FTID=VAR_008155.
FT CONFLICT 1659 1660 LQ -> FE (IN REF. 3).
FT SEQUENCE 1690 AA; 164095 MW; E1E72F283A72BAAE CRC64;
Query Match 33.1%; Score 45; DB 1; Length 1690;
Best Local Similarity 56.2%; Pred. No. 1.2e+02;
Matches 9; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 9 PKGVKHLKDFPILPGE 24
||| : | | | | |
Db 904 PKGPRGLPGPGFPGFE 919

RESULT 15
ID YCF2_PINTH STANDARD; PRT; 2054 AA.
AC P41653;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical 244.6 kDa protein ycf2 (ORF 2054).
GN YCF2.
OS Pinus thunbergii (Green pine) (Japanese black pine).
OG Chloroplast.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Coniferopsida; Coniferales; Pinaceae; Pinus.
OX NCBI_taxid=3350;
RN [1]

RP SEQUENCE FROM N.A.
RX MEDLINE=95024047; PubMed=7937893;
RA Wakasugi T., Tsudzuki J., Ito S., Nakashima K., Tsudzuki T.,
RA Sugiura M.;
RT "Loss of all ndh genes as determined by sequencing the entire
RT chloroplast genome of the black pine Pinus thunbergii.";
RL Proc. Natl. Acad. Sci. U.S.A. 91:9794-9798(1994).
CC -1- FUNCTION: NOT YET KNOWN.
CC -1- SIMILARITY: BELONGS TO THE YCF2 FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; D17510; BAA04460.1; -;
CC InterPro; IPR003959; AAA_ATPase_cent.
CC Pfam; PF00004; AAA; 1.
CC Chloroplast; Hypothetical protein.
KW Chloroplast; Hypothetical protein.
SQ SEQUENCE 2054 AA; 244604 MW; 6F5E92D078E33A9A CRC64;
Query Match 33.1%; Score 45; DB 1; Length 2054;
Best Local Similarity 61.5%; Pred. No. 1.6e+02;
Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 RPNYSRRLPKGVK 13
||| : | | | | |
Db 786 RPNYRNRLPSDMK 798

Search completed: January 23, 2003, 09:01:12
Job time : 14 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 23, 2003, 09:00:58 ; Search time 29 Seconds
(without alignments)
177.627 Million cell updates/sec

Title: SEQ2-ASNAT3

Perfect score: 136

Sequence: 1 PNYSRRLPKGVKHLKDFPILPGEI 25

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_21:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phase:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_rvirus:*
- 16: sp_bacteriap:*
- 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	101	74.3	2343	6	018806
2	101	74.3	2343	6	062730
3	51	37.5	166	4	096AX6
4	51	37.5	1607	10	065506
5	50.5	37.1	589	2	09S158
6	50.5	37.1	935	5	09VE79
7	50	36.8	388	10	08SAW1
8	50	36.8	579	4	096A81
9	50	36.8	706	10	08S5J1
10	50	36.8	958	16	09PEA7
11	50	36.8	2444	5	09VP88
12	50	36.8	2618	5	09NGF9
13	49	36.0	209	11	09B5D5
14	49	36.0	412	17	058994
15	48.5	35.7	386	11	0920V9
16	48.5	35.7	390	11	0920W1

17	48.5	35.7	390	11	Q920W0	mus musculus
18	48.5	35.7	390	11	Q920V8	mus musculus
19	48.5	35.7	390	11	Q920V7	mus musculus
20	48.5	35.7	390	11	Q920V6	mus spicile
21	48.5	35.7	390	11	Q91VB5	mus musculus
22	48.5	35.7	390	11	Q91V20	mus musculus
23	48.5	35.7	711	10	08RYK8	oryza sativ
24	48.5	35.7	1253	10	Q9SCZ3	arabidopsis
25	48	35.3	35	7	Q9MMB4	arabidopsis
26	48	35.3	156	16	Q9HXF9	alcolapia a
27	48	35.3	264	16	Q8YKJ6	pseudomonas
28	48	35.3	360	11	Q922M1	anabaena sp
29	48	35.3	600	11	Q9QYV6	mus musculus
30	48	35.3	600	11	Q8VDR3	rattus norv
31	48	35.3	943	10	Q8VY19	mus musculus
32	48	35.3	1161	10	Q9VLR20	arabidopsis
33	48	35.3	1690	3	Q74835	arabidopsis
34	47.5	34.9	350	2	Q93F54	schizosacch
35	47.5	34.9	352	2	Q93F60	pseudomonas
36	47.5	34.9	352	2	Q93F48	pseudomonas
37	47.5	34.9	401	11	Q9JIK2	pseudomonas
38	47.5	34.9	433	11	Q99N88	cricketul
39	47	34.6	187	16	Q8ZMP8	rattus norv
40	47	34.6	444	10	Q43756	salmonella
41	47	34.6	502	5	Q9N412	caenorhabdi
42	47	34.6	615	10	Q9FYG4	arabidopsis
43	47	34.6	620	10	Q9CAF8	arabidopsis
44	47	34.6	803	10	Q9FXC8	arabidopsis
45	47	34.6	826	4	Q9HAS8	homo sapien

ALIGNMENTS

RESULT 1

O18806 O18806 PRELIMINARY; PRT; 2343 AA.
ID O18806;
DC O18806;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE Factor VIII.
GN F8.
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OC NCBI_TaxID=9615;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LIVER;
RA Cameron C., Nottley C., Hoyle S., McGlynn L., Hough C., Kamisue S.,
RA Giles A., Lilliecrap D.;
RL "The canine factor VIII cDNA and 5' flanking sequence.";
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS.
DR EMBL; AF016234; AAB87412.1; -;
DR HSSP; P00451; ICFG.
DR InterPro; IPR001117; Cu-oxidase.
DR InterPro; IPR000421; FA58_C.
DR Pfam; PF00394; Cu-oxidase; 3.
DR Pfam; PF00754; F5_F8_type_C; 2.
DR SMART; SM00231; FA58C; 2.
DR PROSITE; PS01285; FA58C_1; 2.
DR PROSITE; PS01286; FA58C_2; 2.
DR PROSITE; PS00079; MULTICOPPER_OXIDASE1; 3.
SQ SEQUENCE 2343 AA; 265829 MW; A854FAE571C3B399 CRC64;

Query Match 74.3%; Score 101; DB 6; Length 2343;
Best Local Similarity 79.2%; Pred. No. 2.4e-06;
Matches 19; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 2 PNYSRRLPKGVKHLKDFPILPGEI 25
| :: ||||| ||||| |||||

121 PAYLPRIPPGKVRRAFCIKKLENFVPGPG 150

OS Comamonas testosteroni (Pseudomonas testosteroni)
OC Bacteria: proteobacteria: beta subdivision: Comamonas testosteroni (Pseudomonas testosteroni)

OS Comamonas testosteroni (Pseudomonas testosteroni).
OC Bacteria: proteobacteria: beta subdivision: Comamonadaceae: Comamonas.

OX NCBI_TaxID=285;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=TA441;
 RX MEDLINE=20035595; PubMed=10537203;
 RA Arai H., Yamamoto T., Ohishi T., Shimizu T., Nakata T., Kudo T.;
 RT Genetic organization and characterization of the 3 (3'-
 RT hydroxyphenyl)propionic acid degradation pathway of *Comamonas*
 RT testosteroni TA441.;
 RL Microbiology 145:2813-2820[1999].
 CC -!- COFACTOR: FAD (BY SIMILARITY).
 DR EMBL; AB024335; BAA82878.1; -;
 DR InterPro: IPR001327; FAD_Pyr_redox.
 DR InterPro: IPR000733; Flav_monooxygenase.
 DR InterPro: IPR002114; HPr_SerF_site.
 DR InterPro: IPR002938; MoxF_FAD_binding.
 DR InterPro: IPR000205; NAD_binding.
 DR InterPro: IPR001013; Pyridine_redox_2.
 DR InterPro: IPR003042; Rng_monooxygenase.
 DR InterPro: IPR000594; ThiF_domain.
 DR Pfam: PF01494; FAD_binding_3; 1.
 DR Pfam: PF01360; Monooxygenase; 1.
 DR PRINTS: PR00368; FADPNR.
 DR PRINTS: PR00469; PNDRTASE1.
 DR PRINTS: PR00420; RNCMNKGNAE.
 DR PROSITE: PS00589; PTS_HPR_SER; UNKNOWN_1.
 KW FAD; Flavoprotein; Oxidoreductase.
 SQ SEQUENCE 589 AA; 65297 MW; 7C4CF905825CEB744 CRC64;

 Query Match 37.1%; Score 50.5; DB 2; Length 589;
 Best Local Similarity 41.7%; Pred. No. 29;
 Matches 10; Conservative 5; Mismatches 8; Indels 1; Gaps 1;

 QY 1 RPNYSRRRLPGVKVHLKDPFLPGE 24
 DB 218 RPYVSAALPHGIRRE-EFVMPGE 240

 RESULT 6
 QYVE79 PRELIMINARY; PRT; 935 AA.
 AC QYVE79;
 DT 01-MAY-2000 (TrEMBLrel. 13; Created)
 DT 01-MAY-2000 (TrEMBLrel. 13; Last sequence update)
 DT 01-MAY-2000 (TrEMBLrel. 13; Last annotation update)
 DE CG14309 protein.
 GN CG14309
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BERKELEY;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celisner S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Ananides P.G., Scher S.E., Li P.W., Hoskins R.A., Galie R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abail J.F., Agayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballwee R.M., Basu P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Beeson K.Y., Benos P.V., Berni J., Brockstein P., Brottier P.,
 RA Borkova D., Botchan M.R., Bouck J., Brockstein P., Center A., Chandra I.,
 RA Burtis K.C., Busam D.A., Butler H., Cadenhead L.B., Davies P.,
 RA Cherry J.M., Cavley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Paulis B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes N., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,

RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Iel Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulyov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Rollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Klamos I., Slapston M., Strong R., Sun E.,
 RA Spier E., Spradling A.C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao O., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of *Drosophila melanogaster*."
 RL Science 287:2185-2195(2000).
 DR EMBL; AE003721; AAF55548.1; -;
 DR FlyBase; FBgn0038611; CG14309.
 SQ SEQUENCE 935 AA; 106883 MW; 10E46B721CE5DB3F CRC64;

 Query Match 37.1%; Score 50.5; DB 5; Length 935;
 Best Local Similarity 45.8%; Pred. No. 48;
 Matches 11; Conservative 4; Mismatches 8; Indels 1; Gaps 1;

 QY 2 PNYSRRLPGVKVHLKDPFLPGEI 25
 DB 575 PNYT-RLPEGVVHLKHIEVDSGV 597

 RESULT 7
 QYSAW1 PRELIMINARY; PRT; 388 AA.
 AC QYSAW1;
 DT 01-JUN-2002 (TrEMBLrel. 21; Created)
 DT 01-JUN-2002 (TrEMBLrel. 21; Last sequence update)
 DT 01-JUN-2002 (TrEMBLrel. 21; Last annotation update)
 DE Putative wall-associated protein kinase.
 GN OSJNBA0051J07.1.
 OS Oryza sativa (Rice).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 OC Ehrhartoideae; Oryzaceae; Oryza.
 OX NCBI_TaxID=4530;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Wang R.A., Yu Y., Soderlund C., Chen M., Kim H.-R., Rambo T.,
 RA Sasaki C., Henry D., Oates R., Simmons J.;
 RT "Rice Genomic Sequence."
 RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AC098586; AAL77114.1; -;
 KW Kinase.
 SQ SEQUENCE 388 AA; 42179 MW; 0F4F47E6F087FF91 CRC64;

 Query Match 36.8%; Score 50; DB 10; Length 388;
 Best Local Similarity 47.6%; Pred. No. 22;
 Matches 10; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

 QY 3 NYSRRRLPGVKVHLKDPFLPGE 23
 DB 240 NFSKYPKGVPLVDFAIROG 260

 RESULT 8
 QYSAW1 PRELIMINARY; PRT; 579 AA.
 AC QYSAW1;
 DT 01-DEC-2001 (TrEMBLrel. 19; Created)
 DT 01-DEC-2001 (TrEMBLrel. 19; Last sequence update)

DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE CDNA FLJ131952 fis. clone NT2P7007221, weakly similar to Rattus
DE norvegicus schlafon-4 (SLFN-4) mRNA.
OS Homo sapiens (human)
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Nimomiya K., Wagatsuma M., Kanda K., Kondo H., Yokoi T., Kodaira H.,
RA Furuya T., Takahashi M., Kikkawa E., Omura Y., Abe K., Kamihara K.,
RA Katsuta N., Sato K., Tanikawa M., Yamazaki M., Sugiyama T., Irie R.,
RA Otsuki T., Sato H., Wakamatsu A., Ishii S., Yanamoto J., Isono Y.,
RA Kawai-Hio Y., Saito K., Nishikawa T., Kimura K., Yamashita H.,
RA Matsuo K., Nakamura Y., Sakine M., Kikuchi H., Murakawa K.,
RA Kanehori K., Takahashi-Fujii A., Oshima A., Sugiyama A., Kawakami B.,
RA Suzuki Y., Sugano S., Nagahara K., Masuho Y., Nagai K., Isogai T.,
RT "NEDO human cDNA sequencing project."
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK056514; BAB71201.1;
SQ SEQUENCE 579 AA; 65949 MW; 195AEFB2F4235C84 CRC64;
Query Match 36.8%; Score 50; DB 4; Length 579;
Best Local Similarity 40.0%; Pred. No. 34;
Matches 12; Conservative 5; Mismatches 5; Indels 8; Gaps 2;
QY 1 RPNTSRRLPGVKHLKD----FPILPGEI 25
DB 61 RPVTYSK---KGLHRADLQHLFPVPSGL 87
RESULT 9
ID Q855J1 PRELIMINARY; PRT; 706 AA.
AC Q855J1
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DE 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DE 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Putative wall-associated protein kinase.
GN OSJNB40011A24.33.
OS Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzeae; Oryza.
OX NCBI_TaxID=39947;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN-CV. NIPPONBARE;
RA McCombie W.R., de la Bastide M., Spiegel L., Preston R., Ferraro K.,
RA Kuit K., Nascimento L., Zutavern T., Balija V., Bell M., Baker J.,
RA Santos L., Miller B., Katzenberger F., Muller S., King L., Yang C.,
RA Dike S., O'Shaughnessy A., Palmer L., Dedhia N.;
RT "Genomic sequence for Oryza sativa, Nipponbare strain, clone
RT OSJNB40011A24, from chromosome 10, complete sequence."
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC113336; NAM01181.1;
KW Kinase
SQ SEQUENCE 706 AA; 77555 MW; 91B6F07E9B90BF8 CRC64;
Query Match 35.8%; Score 50; DB 10; Length 706;
Best Local Similarity 47.6%; Pred. No. 42;
Matches 10; Conservative 4; Mismatches 7; Indels 0; Gaps 0;
QY 3 NYSRRLPGVKHLKDFPILPG 23
DB 240 NFSKKYKGVPLVIDFAIROG 260
RESULT 10
ID Q9PBA7 PRELIMINARY; PRT; 958 AA.
AC Q9PBA7
DT 01-OCT-2000 (TrEMBLrel. 15, Created)

DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DE 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Hypothetical Protein Xf2237.
GN Xf2237.
OS Xylella fastidiosa.
OC Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;
OC Xylella.
OX NCBI_TaxID=2371;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-9A5C;
RX MEDLINE=20365717; PubMed=10910347;
RA Simpson A.J.G., Reinsch F.C., Arruda P., Abreu F.A., Acencio M.,
RA Alvarenga R., Alves L.M.C., Araya J.E., Baia G.S., Baptista C.S.,
RA Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,
RA Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carer H.,
RA Coutinho N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,
RA Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorfy H.,
RA Facincani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,
RA Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furlan L.R.,
RA Garnier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,
RA Ho P.L., Hoheisel J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,
RA Krieger J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite L.C.C.,
RA Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,
RA Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Marino C.L.,
RA Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,
RA Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,
RA Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.F.S.,
RA Nhani A., Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,
RA de Oliveira W.C., de Oliveira R.C., Palmieri O.A., Paris A.,
RA Peixoto B.R., Pereira G.A.G., Pereira H.A., Jr., Pesquero J.B.,
RA Queglio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,
RA de Rosa V.E., Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,
RA da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A., Jr.,
RA da Silveira J.F., Silvestri M.L.Z., Silveira W.J., de Souza A.A.,
RA de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tshako M.H.,
RA Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,
RA Zago M.A., Zatz M., Meidanis J., Setubal J.C.;
RT "The genome sequence of the plant pathogen Xylella fastidiosa."
RL Nature 406:151-159(2000).
DR EMBL; AE004036; AAF85036.1;
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 958 AA; 103030 MW; 4EF720E0D0333EDB CRC64;
Query Match 36.8%; Score 50; DB 16; Length 958;
Best Local Similarity 47.1%; Pred. No. 59;
Matches 8; Conservative 4; Mismatches 5; Indels 0; Gaps 0;
QY 6 RRLPGVKHLKDFPILP 22
DB 124 NRIPKGIQETPDITIP 140
RESULT 11
ID Q9VP88 PRELIMINARY; PRT; 2444 AA.
AC Q9VP88
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DE 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DE 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE Putative PAP protein (CG9936 protein).
GN PAP OR CG9936.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,

```

RP SEQUENCE FROM N.A.
RA Naïrz K., Hafen E.;
RR "Isolation of flytrap (pap), the Drosophila TRAP240 homologue.";
RN Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
RL [3]
RN SEQUENCE FROM N.A.
RA Boube M., Faucher C., Joulia L., Cribbs D.L., Bourbon H.M.;
RR "The Drosophila Mediator proteins DTRAP240 and DTRAP80 are
RN transcriptional cofactors of the proboscipedia and Sex combs reduced
RL homeotic proteins.";
RN Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
RL [4]
RN SEQUENCE FROM N.A.
RA Treisman J.E.;
RR "Drosophila homologs of the transcriptional coactivation complex
RN subunits TRAP240 and TRAP230 are required for identical processes in
RL eye-antennal disc development.";
RL Development 0:0-0(2001).
DR EMBL; AF227215; AAF43172.1; -
DR EMBL; AF226855; AAF36691.1; -
DR EMBL; AF227214; AAF43021.1; -
DR EMBL; AF324425; AAG48327.1; -
DR FlyBase; FBgn0024200; pap.
RN SEQUENCE 2618 AA; 280021 MW; 735A8A502076844E CRC64;
LSQ
Query Match 36.8%; Score 50; DB 5; Length 2618;
Best Local Similarity 40.9%; Pred. No. 1.8e-02;
Matches 9; Conservative 6; Mismatches 5; Indels 2; Gaps
OY 1 RPNYSRLPKGVKHLKDFPILP 22
||| : : : : : : : :
DB 836 RPYAKNLYEGONHVK--PVWP 855
||| : : : : : : : :
RESULT 13
Q9D5D5
ID Q9D5D5 PRELIMINARY; PRT; 209 AA.
AC Q9D5D5;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)

```

DE 4930455806Rik protein.
GN 4930455806RIK.
OS MUS MUSCULUS (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]

MEDLINE:21985660, PubMed:11217851.
 RA Kawai J, Shinagawa A, Shibata K, Yoshino M, Itoh M, Ishii Y,
 RA Arakawa T, Hara A, Fukunishi Y, Konno H, Adachi J, Fukuda S,
 RA Aizawa K, Izawa M, Nishii K, Kiyosawa H, Konno S, Yamaneke I,
 RA Saito C, Okazaki Y, Gotohori T, Sone H, Kanokawa T, Saito R,
 RA Kadote K, Matsuda H, A. Ashburner J, Batazou S, Casavant T,
 RA Fleischmann W, Gaasterland T, Nikaido I, Kresole G, Quackenbush J,
 RA Ruehl P, Lewis S, Matsuo T, Kunita K, Komita H, Wagner D, Washio T,
 RA Sakai K, Okido T, Furuno M, Aono H, Baldarelli R, Barsh G,
 RA Blake J, Bocelli D, Bojunga N, Carninci P, de Bonaldo M F,
 RA Brownstein M J, Bult C, Fletcher C, Fujita M, Gariboldi M,
 RA Gustincich S, Hill B, Hofmann M, Hume D.A, Kamiya M, Lee N.H.,
 RA Lyons P, Marchionni L, Mashima J, Mazzarelli J, Momberto P.,
 RA Nordone P, Ring B, Ringwald M, Rodriguez I., Sakamoto N,
 RA Sasaki H, Sato K, Schoenbach C, Seya T, Shibata Y, Storch K.F.,
 RA Suzuki H, Toyo-oka K, Wang K.H., Weiz C., Whitaker C., Wilming L,
 RA Wysshaw-Boris A, Yoshida K, Hasegawa Y, Kawaji H, Kohsaki S.,
 RA Hayashizaki Y;
 RA "Functional annotation of a full-length mouse cDNA collection.";
 RA Nature 409:685-690(2001).
 RA EMBL: AK015457; BAB29855.1; -
 RA MGD: MGI:1921905; 4930455B046Rik.

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 23, 2003, 09:01:58 ; Search time 14 Seconds
(without alignments)
52.541 Million cell updates/sec

Title: SEQ2-ASNAT3
Perfect score: 136
Sequence: 1 RPNYSRRLLPKGVKHLKDFPILPGEI 25

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA: *
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep: *
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep: *
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep: *
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep: *
5: /cgn2_6/ptodata/1/iaa/PTUS_COMB.pep: *
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	127	93.4	1438	4	US-09-209-916-1 Sequence 1, Appli
2	127	93.4	1471	1	US-08-683-839B-3 Sequence 3, Appli
3	127	93.4	1661	2	US-08-882-083-2 Sequence 2, Appli
4	127	93.4	1661	2	US-08-558-107-2 Sequence 2, Appli
5	127	93.4	1661	4	US-09-243-539-2 Sequence 2, Appli
6	127	93.4	2332	1	US-07-864-004B-4 Sequence 4, Appli
7	127	93.4	2332	1	US-08-251-937A-4 Sequence 4, Appli
8	127	93.4	2332	1	US-08-212-133A-2 Sequence 2, Appli
9	127	93.4	2332	1	US-08-276-594A-2 Sequence 2, Appli
10	127	93.4	2332	1	US-08-474-503-2 Sequence 2, Appli
11	127	93.4	2332	2	US-08-670-707A-2 Sequence 2, Appli
12	127	93.4	2332	4	US-09-037-601-2 Sequence 3, Appli
13	127	93.4	2332	4	US-09-324-867-3 Sequence 2, Appli
14	127	93.4	2332	4	US-09-315-179-2 Sequence 2, Appli
15	127	93.4	2332	4	US-09-523-656-2 Sequence 2, Appli
16	127	93.4	2332	5	PCT-US93-03275-4 Sequence 4, Appli
17	127	93.4	2332	5	PCT-US94-13200-2 Sequence 2, Appli
18	127	93.4	2351	1	US-08-121-202-2 Sequence 2, Appli
19	127	93.4	2351	1	US-08-366-851A-2 Patent No. 5171844
20	127	93.4	2351	6	US-07-1844-2 Patent No. 5422260
21	127	93.4	2351	6	US-09-324-867-2 Sequence 2, Appli
22	101	74.3	2343	4	US-09-324-867-4 Sequence 4, Appli
23	94	69.1	2304	1	US-08-212-133A-8 Sequence 8, Appli
24	94	69.1	2319	1	US-08-474-503-6 Sequence 6, Appli
25	94	69.1	2319	1	US-08-670-707A-6 Sequence 6, Appli
26	94	69.1	2319	4	US-09-037-601-6 Sequence 6, Appli
27	94	69.1	2319	4	US-09-523-656-38 Sequence 38, Appli

28	94	69.1	2319	4	US-09-315-179-6 Sequence 6, Appli
29	94	69.1	2319	4	US-09-523-656-28 Sequence 28, Appli
30	94	69.1	2319	5	PCT-US94-13200-6 Sequence 6, Appli
31	75	55.1	367	1	US-07-864-004B-2 Sequence 2, Appli
32	75	55.1	367	1	US-08-251-937A-2 Sequence 2, Appli
33	75	55.1	367	5	PCT-US93-03275-2 Sequence 2, Appli
34	75	55.1	368	1	US-08-212-133A-6 Sequence 6, Appli
35	75	55.1	368	1	US-08-474-503-4 Sequence 4, Appli
36	75	55.1	368	2	US-08-670-707A-4 Sequence 4, Appli
37	75	55.1	368	4	US-09-037-601-4 Sequence 4, Appli
38	75	55.1	368	4	US-09-315-179-4 Sequence 4, Appli
39	75	55.1	368	4	US-09-523-656-4 Sequence 4, Appli
40	75	55.1	368	5	PCT-US94-13200-4 Sequence 4, Appli
41	75	55.1	541	1	US-08-121-202-4 Sequence 39, Appli
42	75	55.1	1443	2	US-08-670-707A-39 Sequence 39, Appli
43	75	55.1	1443	4	US-09-037-601-39 Sequence 39, Appli
44	75	55.1	1443	4	US-09-315-179-39 Sequence 39, Appli
45	75	55.1	1467	4	US-09-523-656-38 Sequence 38, Appli

ALIGNMENTS

RESULT 1
US-09-209-916-1
; Sequence 1, Application US/09209916
; Patent No. 6358703
; GENERAL INFORMATION:
; APPLICANT: Cho, Myung-Sam
; APPLICANT: Chan, Sham-Yuen
; APPLICANT: Kelsey, William
; APPLICANT: Yee, Helena
; TITLE OF INVENTION: Expression System for Factor VIII
; FILE REFERENCE: MSB-7255
; CURRENT APPLICATION NUMBER: US/09/209,916
; CURRENT FILING DATE: 1998-12-10
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 1438
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Derived from
; OTHER INFORMATION: human factor VIII sequence
US-09-209-916-1

Query Match 93.4%; Score 127; DB 4; Length 1438;
Best Local Similarity 96.0%; Pred. No. 3.3e-11;
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 RPNYSRRLLPKGVKHLKDFPILPGEI 25
|||
Db 484 RPLYSRRLLPKGVKHLKDFPILPGEI 508

RESULT 2

US-08-683-839B-3
; Sequence 3, Application US/08683839B
; Patent No. 5744326
; GENERAL INFORMATION:
; APPLICANT: Irl, Charles . R. et al.
; TITLE OF INVENTION: Use of Viral Cis-Acting Post-Transcriptional
; TITLE OF INVENTION: Regulatory Sequences To Increase Expression of
; TITLE OF INVENTION: Ironless Genes Containing Near-Consensus Splice Sites
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 60 State Street, suite 510
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109-1875

```

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/683,839B
; FILING DATE: 11-MARCH-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Remillard, Jane E.
; REGISTRATION NUMBER: 38,872
; REFERENCE/DOCKET NUMBER: TTI-138
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)227-7400
; TELEFAX: (617)227-5941
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1471 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-683-839B-3

```

```

Query Match          93.4%; Score 127; DB 1; Length 1471;
Best Local Similarity 96.0%; Pred. No. 3.4e-11;
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

QY 1 RPNYSRRLPKGVKHLKDFPILPGEI 25
   |||||||||||||||||||
DB 503 RPLYSRRLPKGVKHLKDFPILPGEI 527

```

```

RESULT 3
US-08-882-083-2
; Sequence 2, Application US/08882083
; Patent No. 5869292
; GENERAL INFORMATION:
; APPLICANT: VOORBERG, Johannes J.
; TITLE OF INVENTION: HYBRID PROTEINS WITH MODIFIED ACTIVITY
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite 500
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/882,083
; FILING DATE:
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/558,107
; FILING DATE: 13-NOV-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: ISACSON, John P.
; REGISTRATION NUMBER: 33,715
; REFERENCE/DOCKET NUMBER: 30472/212
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)672-5300
; TELEFAX: (202)672-5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:

```

```

; LENGTH: 1661 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-882-083-2
; Query Match          93.4%; Score 127; DB 2; Length 1661;
; Best Local Similarity 96.0%; Pred. No. 3.9e-11;
; Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

QY 1 RPNYSRRLPKGVKHLKDFPILPGEI 25
   |||||||||||||||||||
DB 503 RPLYSRRLPKGVKHLKDFPILPGEI 527

```

```

RESULT 4
US-08-558-107-2
; Sequence 2, Application US/08558107
; Patent No. 5910481
; GENERAL INFORMATION:
; APPLICANT: VOORBERG, Johannes J.
; TITLE OF INVENTION: HYBRID PROTEINS WITH MODIFIED ACTIVITY
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite 500
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/558,107
; FILING DATE: 13-NOV-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: ISACSON, John P.
; REGISTRATION NUMBER: 33,715
; REFERENCE/DOCKET NUMBER: 30472/212
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)672-5300
; TELEFAX: (202)672-5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1661 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-558-107-2

```

```

Query Match          93.4%; Score 127; DB 2; Length 1661;
Best Local Similarity 96.0%; Pred. No. 3.9e-11;
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

QY 1 RPNYSRRLPKGVKHLKDFPILPGEI 25
   |||||||||||||||||||
DB 503 RPLYSRRLPKGVKHLKDFPILPGEI 527

```

```

RESULT 5
US-09-243-539-2
; Sequence 2, Application US/09243539
; Patent No. 6130203
; GENERAL INFORMATION:
; APPLICANT: VOORBERG, Johannes J.
; TITLE OF INVENTION: HYBRID PROTEINS WITH MODIFIED ACTIVITY
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:

```

```

1  RESULT 6
2  US-07-864-004B-4
3  ; Sequence 4, Application US/07864004B
4  ; Patent No. 5364771
5  ; GENERAL INFORMATION:
6  ; APPLICANT: Lollar, John S.
7  ; APPLICANT: Runge, Marschall S.
8  ; TITLE OF INVENTION: Hybrid Human/Porcine Factor VIII
9  ; NUMBER OF SEQUENCES: 6
10 ; CORRESPONDENCE ADDRESS:
11 ; ADDRESSEE: Kilpatrick & Cody
12 ; STREET: 1100 Peachtree Street
13 ; CITY: Atlanta
14 ; STATE: Georgia
15 ; COUNTRY: US
16 ; ZIP: 30309
17 ; COMPUTER READABLE FORM:
18 ; MEDIUM TYPE: Floppy disk
19 ; COMPUTER: IBM PC compatible
20 ; OPERATING SYSTEM: PC-DOS/MS-DOS
21 ; SOFTWARE: PatentIn Release #1.0, Version #1.25
22 ; CURRENT APPLICATION DATA:
23 ; APPLICATION NUMBER: US/07864, 004B
24 ; FILING DATE: 07 APRIL 1992
25 ; CLASSIFICATION: 435
26 ; ATTORNEY/AGENT INFORMATION:
27 ; NAME: Pabst, Patrea L.
28 ; REGISTRATION NUMBER: 31,284
29 ; REFERENCE/DOCKET NUMBER: EMU106
30 ; TELECOMMUNICATION INFORMATION:

```

```

1  TITLE OF INVENTION: Hybrid Human/Porcine Factor VIII
2  NUMBER OF SEQUENCES: 10
3  CORRESPONDENCE ADDRESS:
4  ADDRESSEE: Kilpatrick & Cody
5  STREET: 1100 Peachtree Street
6  CITY: Atlanta
7  STATE: Georgia
8  COUNTRY: US
9  ZIP: 30309
10
11 COMPUTER READABLE FORM:
12 MEDIUM TYPE: Floppy disk
13 COMPUTER: IBM PC compatible
14 OPERATING SYSTEM: PC-DOS/MS-DOS
15 SOFTWARE: PatentIn Release #1.0, Version #1.25
16
17 CURRENT APPLICATION DATA:
18 APPLICATION NUMBER: US/08/251,937A
19 FILING DATE: 31-MAY-1994
20 CLASSIFICATION: 435
21 PRIOR APPLICATION DATA:
22 APPLICATION NUMBER: US 07/864,004
23 FILING DATE: 07-APR-1992
24 ATTORNEY/AGENT INFORMATION:
25 NAME: Pratt, John S.
26 REGISTRATION NUMBER: 29,476
27 REFERENCE/DOCKET NUMBER: ERMU06DIV
28 TELECOMMUNICATION INFORMATION:
29 TELEPHONE: 404-815-6367
30 TELEFAX: 404-815-6555
31
32 INFORMATION FOR SEQ ID NO: 4:
33 SEQUENCE CHARACTERISTICS:
34 LENGTH: 2332 amino acids
35 TYPE: amino acid
36 STRANDEDNESS: single
37 TOPOLOGY: linear
38 MOLECULE TYPE: protein
39 HYPOTHETICAL: NO
40 ANTI-SENSE: NO
41 FRAGMENT TYPE: N-terminal
42 ORIGINAL SOURCE:
43 ORGANISM: Homo sapien

```

RESULT 9
US-08-276-594A-2


```
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/474,503
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Pratt, John S.
; REGISTRATION NUMBER: 29,476
; REFERENCE/DOCKET NUMBER: EMU106CIP(3)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 404-815-6500
; TELEFAX: 404-815-6555
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2332 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: YES
; ANTI-SENSE: NO
; FRAGMENT TYPE: N-terminal
; ORIGINAL SOURCE:
; ORGANISM: Homo sapien
; TISSUE TYPE: Liver cDNA sequence
; US-08-474-503-2
```

```
Query Match 93.4%; Score 127; DB 1; Length 2332;
Best Local Similarity 96.0%; Pred. No. 5.7e-11;
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
QY 1 RPNYSRRLPKGVKHLKDPILPGEI 25
|| |||||
Db 484 RPLYSRRLPKGVKHLKDPILPGEI 508
```

```
RESULT 11
US-08-670-707A-2
; Sequence 2, Application US/08670707A
; Patent No. 5859204
; GENERAL INFORMATION:
; APPLICANT: Lollar, John S.
; TITLE OF INVENTION: Hybrid Human/Animal Factor VIII
; NUMBER OF SEQUENCES: 40
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Greenlee, Winner and Sullivan, P.C.
; STREET: 5370 Manhattan Circle Suite 201
; CITY: Boulder
; STATE: Colorado
; COUNTRY: USA
; ZIP: 80303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/670,707A
; FILING DATE: 26-JUN-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/US94/13200
; FILING DATE: 15-NOV-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/212,133
; FILING DATE: 11-MAR-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/864,004
; FILING DATE: 07-APR-1992
; ATTORNEY/AGENT INFORMATION:
```

```
; NAME: Greenlee, Lorraine L.
; REGISTRATION NUMBER: 27,894
; REFERENCE/DOCKET NUMBER: 75-95F
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 303/499-8080
; TELEFAX: 303/499-8089
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2332 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: not relevant
; MOLECULE TYPE: protein
; HYPOTHETICAL: YES
; ANTI-SENSE: NO
; FRAGMENT TYPE: N-terminal
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; TISSUE TYPE: Liver
; US-08-670-707A-2
```

```
Query Match 93.4%; Score 127; DB 2; Length 2332;
Best Local Similarity 96.0%; Pred. No. 5.7e-11;
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
QY 1 RPNYSRRLPKGVKHLKDPILPGEI 25
|| |||||
Db 484 RPLYSRRLPKGVKHLKDPILPGEI 508
```

```
RESULT 12
US-09-037-601-2
; Sequence 2, Application US/09037601
; Patent No. 6180371
; GENERAL INFORMATION:
; APPLICANT: Lollar, John S.
; TITLE OF INVENTION: Hybrid Human/Animal Factor VIII
; NUMBER OF SEQUENCES: 40
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Greenlee, Winner and Sullivan, P.C.
; STREET: 5370 Manhattan Circle Suite 201
; CITY: Boulder
; STATE: Colorado
; COUNTRY: USA
; ZIP: 80303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/037,601
; FILING DATE: 26-JUN-1996
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/US94/13200
; FILING DATE: 15-NOV-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/212,133
; FILING DATE: 11-MAR-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/864,004
; FILING DATE: 07-APR-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Ferber, Donna M.
; REGISTRATION NUMBER: 33,878
; REFERENCE/DOCKET NUMBER: 75-95F
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 303/499-8080
; TELEFAX: 303/499-8089
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2332 amino acids
```

```

1 GENERAL INFORMATION:
2 APPLICANT: Lollar, John S
3 TITLE OF INVENTION: Modified Factor VIII
4 FILE REFERENCE: 75-95H
5 CURRENT APPLICATION NUMBER: US/09/315,179
6 CURRENT FILING DATE: 1999-05-20
7 EARLIER APPLICATION NUMBER: U.S. 09/037,600
8 EARLIER FILING DATE: 1998-03-10
9 EARLIER APPLICATION NUMBER: U.S. 08/670,700
10 EARLIER FILING DATE: 1996-06-26
11 EARLIER APPLICATION NUMBER: PCT/US97/111555
12 EARLIER FILING DATE: 1997-06-26
13 EARLIER APPLICATION NUMBER: PCT/US94/132000

```

Qy 1 RPNYSRRLPKGVKHLKDFPILPGEI 25
 ++
 Db 484 RPLYSRRLPKGVKHLKDFPILPGEI 508

Search completed: January 23, 2003, 09:04:50
Job time : 17 secs

; TITLE OF INVENTION: Expression System for Factor VIII
; FILE REFERENCE: MSB-7255-2
; CURRENT APPLICATION NUMBER: US/10/047,257
; CURRENT FILING DATE: 2002-01-15
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 1438
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Derived from
; OTHER INFORMATION: human factor VIII sequence
US-10-047-257-1

Query Match 93.4%; Score 127; DB 12; Length 1438;
Best Local Similarity 96.0%; Pred. No. 2.8e-11;
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 RPNYSRRLPKGVKHLKDFPILPGEI 25
|| |||||
Db 484 RPLYSRRLPKGVKHLKDFPILPGEI 508

RESULT 3

US-10-095-718-2
; Sequence 2, Application US/10095718
; Patent No. US20020131956A1
; GENERAL INFORMATION:
; APPLICANT: Walsh, Christopher
; APPLICANT: Chao, Hengjun
; APPLICANT: Burstein, Haim
; APPLICANT: Lynch, Carmel
; APPLICANT: Stepan, Tony
; APPLICANT: Munson, Keith
; TITLE OF INVENTION: Adeno-Associated Virus Vectors Encoding Factor VIII and
; FILE REFERENCE: 35052/204375
; CURRENT APPLICATION NUMBER: US/10/095,718
; CURRENT FILING DATE: 2002-03-12
; PRIOR APPLICATION NUMBER: 09/689,430
; PRIOR FILING DATE: 2001-08-22
; PRIOR APPLICATION NUMBER: 60/158,780
; PRIOR FILING DATE: 1999-10-12
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 1471
; TYPE: PRT
; ORGANISM: Homo sapiens B-domain deleted factor VIII
; FEATURE:
; OTHER INFORMATION: Homo sapiens BDD FVIII
US-10-095-718-2

Query Match 93.4%; Score 127; DB 12; Length 1471;
Best Local Similarity 96.0%; Pred. No. 2.8e-11;
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 RPNYSRRLPKGVKHLKDFPILPGEI 25
|| |||||
Db 503 RPLYSRRLPKGVKHLKDFPILPGEI 527

RESULT 4

US-09-957-641-2
; Sequence 2, Application US/09957641
; Publication No. US20020182670A1
; GENERAL INFORMATION:
; APPLICANT: Emory University
; TITLE OF INVENTION: MODIFIED FACTOR VIII
; FILE REFERENCE: 75-00
; CURRENT APPLICATION NUMBER: US/09/957,641
; CURRENT FILING DATE: 2001-09-16

; PRIOR APPLICATION NUMBER: US 60/234047
; PRIOR FILING DATE: 2000-09-19
; CURRENT APPLICATION NUMBER: US 60/236460
; PRIOR FILING DATE: 2000-09-29
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 2332
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-957-641-2

Query Match 93.4%; Score 127; DB 9; Length 2332;
Best Local Similarity 96.0%; Pred. No. 4.8e-11;
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 RPNYSRRLPKGVKHLKDFPILPGEI 25
|| |||||
Db 484 RPLYSRRLPKGVKHLKDFPILPGEI 508

RESULT 5

US-10-095-718-4
; Sequence 4, Application US/10095718
; Patent No. US20020131956A1
; GENERAL INFORMATION:
; APPLICANT: Walsh, Christopher
; APPLICANT: Chao, Hengjun
; APPLICANT: Burstein, Haim
; APPLICANT: Lynch, Carmel
; APPLICANT: Stepan, Tony
; APPLICANT: Munson, Keith
; TITLE OF INVENTION: Adeno-Associated Virus Vectors Encoding Factor VIII and
; FILE REFERENCE: 35052/204375
; CURRENT APPLICATION NUMBER: US/10/095,718
; CURRENT FILING DATE: 2002-03-12
; PRIOR APPLICATION NUMBER: 09/689,430
; PRIOR FILING DATE: 2001-08-22
; PRIOR APPLICATION NUMBER: 60/158,780
; PRIOR FILING DATE: 1999-10-12
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 1431
; TYPE: PRT
; ORGANISM: canine B-domain deleted factor VIII
US-10-095-718-4

Query Match 74.3%; Score 101; DB 12; Length 1431;
Best Local Similarity 79.2%; Pred. No. 2.6e-07;
Matches 19; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 2 PNYSRRLPKGVKHLKDFPILPGEI 25
| : : |||||
Db 498 PLHTGRPLPKGVKHLKDFPILPGEI 521

RESULT 6

US-09-949-192-49
; Sequence 49, Application US/09949192
; Patent No. US20020142292A1
; GENERAL INFORMATION:
; APPLICANT: Parham, Christi L.
; APPLICANT: Gorman, Daniel L.
; APPLICANT: Kurata, Hirokazu
; APPLICANT: Aral, Naoko
; APPLICANT: Sana, Theodore R.
; APPLICANT: Mattson, Jeanine D.
; APPLICANT: Murphy, Erin E.
; APPLICANT: Savkoor, Chetan
; APPLICANT: Grein, Jeffery
; APPLICANT: Smith, Kathleen M.

APPLICANT: Mcclanahan, Terrill K.
TITLE OF INVENTION: MAMMALIAN GENES; RELATED REAGENTS AND METHODS
FILE REFERENCE: DX01169K
CURRENT APPLICATION NUMBER: US/09/949,192
CURRENT FILING DATE: 2001-09-07
PRIOR APPLICATION NUMBER: 60/231,267
PRIOR FILING DATE: 2000-09-08
NUMBER OF SEQ ID NOS: 53
SOFTWARE: PatentIn version 3.1
SEQ ID NO 49
LENGTH: 897
TYPE: PRT
ORGANISM: Homo sapiens
US-09-949-192-49

Query Match 36.8%; Score 50; DB 10; Length 897;
Best Local Similarity 40.0%; Pred. No. 9.7;
Matches 12; Conservative 5; Mismatches 5; Indels 8; Gaps 2;

QY 1 RPNYSRRLPKGV-KHLKDP 25
DB 379 RPNYSK---KGLEHKADLQQLHFPVPGHL 405

RESULT 7

US-09-863-475A-8
Sequence 8, Application US/09863475A
Patent No. US20020102688A1
GENERAL INFORMATION:

APPLICANT: LOWE, JOHN B.
TITLE OF INVENTION: METHODS AND PRODUCTS FOR THE SYNTHESIS
OF OLIGOSACCHARIDE STRUCTURES ON GLYCOPROTEINS,
GLYCOLIPIDS, OR AS FREE MOLECULES, AND FOR THE ISOLATION
OF CLONED GENETIC SEQUENCES THAT DETERMINE THESE STRUCTURES
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
P.C.
STREET: 1755 Jefferson Davis Highway, Fourth Floor
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22202

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/863,475A

FILING DATE: 24-May-2001

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 07/914,281

FILING DATE: 20-JUL-1992

ATTORNEY/AGENT INFORMATION:

NAME: Lavalleye, Jean-Paul M. P.

REGISTRATION NUMBER: 31,451

REFERENCE/DOCKET NUMBER: 2363-060-55

TELECOMMUNICATION INFORMATION:

TELEPHONE: (703)521-4500

TELEFAX: (703)486-2347

TELEX: 248855 OPAT UR

INFORMATION FOR SEQ ID NO: 8:

SEQUENCE CHARACTERISTICS:

LENGTH: 405 amino acids

TYPE: amino acid

TOPOLOGY: unknown

MOLECULE TYPE: protein

SEQUENCE DESCRIPTION: SEQ ID NO: 8:

US-09-863-475A-8

Query Match 35.7%; Score 48.5; DB 10; Length 405;

Best Local Similarity 50.0%; Pred. No. 6.6;
Matches 10; Conservative 3; Mismatches 6; Indels 1; Gaps 1;

QY 1 RPNYSRRLPKGV-KHLKDP 19
DB 319 RANYERFVRGAFIHVDFF 338

RESULT 8

US-09-733-524-7
Sequence 7, Application US/09733524
Patent No. US20020068347A1
GENERAL INFORMATION:

APPLICANT: The Governors of the University of Alberta, a Canada Corporation
APPLICANT: Taylor, Diane E.
TITLE OF INVENTION: NUCLEIC ACIDS ENCODING ALPHA-1.3
TITLE OF INVENTION: FUCOSYLTRANSFERASES AND EXPRESSION SYSTEMS FOR MAKING AND
TITLE OF INVENTION: EXPRESSING THEM
FILE REFERENCE: 07254/049002
CURRENT APPLICATION NUMBER: US/09/733,524
CURRENT FILING DATE: 2000-12-14
PRIOR APPLICATION NUMBER: 09/092,315
PRIOR FILING DATE: 1998-06-05
PRIOR APPLICATION NUMBER: 60/048,857
PRIOR FILING DATE: 1997-06-06
NUMBER OF SEQ ID NOS: 20
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 7

LENGTH: 432
TYPE: PRT
ORGANISM: Mus musculus
FEATURE:
NAME/KEY: PEPTIDE
LOCATION: (0)...(0)
OTHER INFORMATION: Helicobacter pylori fucosyltransferase - Mfuct4
US-09-733-524-7

Query Match 35.7%; Score 48.5; DB 10; Length 432;

Best Local Similarity 50.0%; Pred. No. 7.1;
Matches 10; Conservative 3; Mismatches 6; Indels 1; Gaps 1;

QY 1 RPNYSRRLPKGV-KHLKDP 19
DB 346 RANYERFVRGAFIHVDFF 365

RESULT 9

US-10-120-319-11
Sequence 11, Application US/10120319
Patent No. US20020164749A1
GENERAL INFORMATION:

APPLICANT: Taylor, Diane E.
TITLE OF INVENTION: ALPHA-1, 3-FUCOSYLTRANSFERASE
FILE REFERENCE: 07254/049001
CURRENT APPLICATION NUMBER: US/10/120,319
CURRENT FILING DATE: 2002-04-09
PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/092,315
PRIOR FILING DATE: EARLIER FILING DATE: 1998-06-05
PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: US 60/048,857
PRIOR FILING DATE: EARLIER FILING DATE: 1997-06-06
NUMBER OF SEQ ID NOS: 22
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 11
LENGTH: 433
TYPE: PRT
ORGANISM: Mus musculus
US-10-120-319-11

Query Match 35.7%; Score 48.5; DB 9; Length 433;

Best Local Similarity 50.0%; Pred. No. 7.1;
Matches 10; Conservative 3; Mismatches 6; Indels 1; Gaps 1;

Query Match 32.7%; Score 44.5; DB 10; Length 376;
Best Local Similarity 52.6%; Pred. No. 25;
Matches 10; Conservative 4; Mismatches 4; Indels 1; Gaps 1;

```

RESULT 13
US-09-893--637-2
; Sequence 2, Application US/09893637
; Patent No. US20020049313A1
; GENERAL INFORMATION:
; APPLICANT: COTTEN, MATTHEW
; BAKER, ADAM
; CHIOCCA, SUSANNA
; TITLE OF INVENTION: Method for Introducing Foreign Material into
; Higher Eukaryotic Cells
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C.
; STREET: 1100 NEW YORK AVE., NW., SUITE 600
; CITY: WASHINGTON
; STATE: DC
; COUNTRY: USA
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/893,637
; FILING DATE: 29-Jun-2001

```

```

; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; FILING DATE: 14-FEB-1997
; APPLICATION NUMBER: PCT/EP95/01989
; FILING DATE: 26-MAY-1995
; APPLICATION NUMBER: DE P 44 18 825.0
; FILING DATE: 30-MAY-1994
; APPLICATION NUMBER: DE P 44 42 587.2
; FILING DATE: 30-NOV-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Peter A. Jackman
; REGISTRATION NUMBER: 45,986
; REFERENCE/DOCKET NUMBER: 0652.1580001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-371-2600
; TELEFAX: 202-371-2540
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 282 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: Protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-09-893-637-2
Query Match 32.4%; Score 44; DB 10; Length 282;
Best Local Similarity 52.6%; Pred. No. 21;
Matches 10; Conservative 2; Mismatches 5; Indels 2; Gaps 1;

OY 4 YSRRLPKGVKHLKD--FPI 20
|| || | : | |||
Db 124 YSDPLPHGITHVGDGLGFI 142

RESULT 14
US-09-970-711-21
; Sequence 21, Application US/09970711
; Patent No. US20020081279A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Adam
; APPLICANT: Cotten, Matthew
; APPLICANT: Chiocca, Susanna
; APPLICANT: Kurzbauer, Robert
; APPLICANT: Schaffner, Gottbold
; TITLE OF INVENTION: Chicken Embryo Lethal Orphan (CELO) Virus
; FILE REFERENCE: 0652-1800001
; CURRENT APPLICATION NUMBER: US/09/970,711
; PRIOR FILING DATE: 2001-10-05
; PRIOR APPLICATION NUMBER: 09/171,461
; PRIOR FILING DATE: 1999-01-12
; PRIOR APPLICATION NUMBER: PCT/EP97/01944
; PRIOR FILING DATE: 1997-04-18
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO: 21
; LENGTH: 282
; TYPE: PRT
; ORGANISM: CELO Virus
; FEATURE:
; OTHER INFORMATION: Position: 37391...38239 /note=ORF8
US-09-970-711-21
Query Match 32.4%; Score 44; DB 10; Length 282;
Best Local Similarity 52.6%; Pred. No. 21;
Matches 10; Conservative 2; Mismatches 5; Indels 2; Gaps 1;

OY 4 YSRRLPKGVKHLKD--FPI 20
|| || | : | |||
Db 124 YSDPLPHGITHVGDGLGFI 142

RESULT 15
US-09-796-858-40
; Sequence 40, Application US/09796858
; Patent No. US20020055139A1
; GENERAL INFORMATION:
; APPLICANT: Holtzmann, Douglas
; TITLE OF INVENTION: NOVEL GENES ENCODING PROTEINS HAVING PROGNOSTIC, DIAGNOSTIC,
; APPLICATION: PREVENTIVE, THERAPEUTIC, AND OTHER USES
; FILE REFERENCE: 7853-226-999
; CURRENT APPLICATION NUMBER: US/09/796,858
; CURRENT FILING DATE: 2001-03-01
; PRIOR APPLICATION NUMBER: 09/223,094
; PRIOR FILING DATE: 1998-12-30
; PRIOR APPLICATION NUMBER: 09/223,546
; PRIOR FILING DATE: 1998-12-30
; PRIOR APPLICATION NUMBER: 09/224,246
; PRIOR FILING DATE: 1998-12-30
; PRIOR APPLICATION NUMBER: 09/312,359
; PRIOR FILING DATE: 1999-05-14
; PRIOR APPLICATION NUMBER: 09/336,536
; PRIOR FILING DATE: 1999-06-18
; PRIOR APPLICATION NUMBER: 09/342,687
; PRIOR FILING DATE: 1999-06-29
; PRIOR APPLICATION NUMBER: 09/399,723
; PRIOR FILING DATE: 1999-09-20
; PRIOR APPLICATION NUMBER: 09/471,179
; PRIOR FILING DATE: 1999-12-23
; PRIOR APPLICATION NUMBER: 09/474,071
; PRIOR FILING DATE: 1999-12-29
; PRIOR APPLICATION NUMBER: 09/474,072
; PRIOR FILING DATE: 1999-12-29
; PRIOR APPLICATION NUMBER: 09/572,002
; PRIOR FILING DATE: 2000-03-14
; PRIOR APPLICATION NUMBER: 09/597,993
; PRIOR FILING DATE: 2000-06-12
; PRIOR APPLICATION NUMBER: 09/599,596
; PRIOR FILING DATE: 2000-06-22
; PRIOR APPLICATION NUMBER: 09/606,565
; PRIOR FILING DATE: 2000-06-29
; PRIOR APPLICATION NUMBER: 09/365,164
; PRIOR FILING DATE: 1999-07-30
; PRIOR APPLICATION NUMBER: 09/630,334
; PRIOR FILING DATE: 2000-07-31
; PRIOR APPLICATION NUMBER: 09/665,666
; PRIOR FILING DATE: 2000-09-20
; NUMBER OF SEQ ID NOS: 50
; SEQ ID NO: 40
; LENGTH: 482
; TYPE: PRT
; ORGANISM: Homo sapiens
; OTHER INFORMATION:
US-09-796-858-40
Query Match 32.4%; Score 44; DB 10; Length 482;
Best Local Similarity 36.4%; Pred. No. 39;
Matches 8; Conservative 4; Mismatches 10; Indels 0; Gaps 0;

OY 1 RPNYSRRRLPKGVKHLKDFPILP 22
|| : | : |||
Db 450 RPSLQKHPSTGGLHAPFPLP 471

Search completed: January 23, 2003, 09:05:29
Job time : 12 secs

```

